Reflect, not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis

HYUNSUNG CHO, Carnegie Mellon University, USA
DAEUN CHOI, KAIST, Republic of Korea
DONGHWI KIM, KAIST, Republic of Korea
WAN JU KANG, KAIST, Republic of Korea
EUN KYOUNG CHOE, University of Maryland, College Park, USA
SUNG-JU LEE, KAIST, Republic of Korea

Digital intervention tools against problematic smartphone usage help users control their consumption on smartphones, for example, by setting a time limit on an app. However, today's social media apps offer a mix of quasiessential and addictive features in an app (e.g., Instagram has following feeds, recommended feeds, stories, and direct messaging features), which makes it hard to apply a uniform logic for all uses of an app without a nuanced understanding of feature-level usage behaviors. We study when and why people regret using different features of social media apps on smartphones. We examine regretful feature uses in four smartphone social media apps (Facebook, Instagram, YouTube, and KakaoTalk) by utilizing feature usage logs, ESM surveys on regretful use collected for a week, and retrospective interviews from 29 Android users. In determining whether a feature use is regretful, users considered different types of rewards they obtained from using a certain feature (i.e., social, informational, personal interests, and entertainment) as well as alternative rewards they could have gained had they not used the smartphone (e.g., productivity). Depending on the types of rewards and the way rewards are presented to users, probabilities to regret vary across features of the same app. We highlight three patterns of features with different characteristics that lead to regretful use. First, "following"-based features (e.g., Facebook's News Feed and Instagram's Following Posts and Stories) induce habitual checking and quickly deplete rewards from app use. Second, recommendation-based features situated close to actively used features (e.g., Instagram's Suggested Posts adjacent to Search) cause habitual feature tour and sidetracking from the original intention of app use. Third, recommendation-based features with bite-sized contents (e.g., Facebook's Watch Videos) induce using "just a bit more," making people fall into prolonged use. We discuss implications of our findings for how social media apps and intervention tools can be designed to reduce regretful use and how feature-level usage information can strengthen self-reflection and behavior changes.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in ubiquitous and mobile computing**; **Smartphones**; *Ubiquitous and mobile computing design and evaluation methods.*

Additional Key Words and Phrases: Smartphone use, digital wellbeing, social media, regret, app feature

Authors' addresses: Hyunsung Cho, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, hyunsung@andrew.cmu.edu; DaEun Choi, KAIST, Daejeon, Republic of Korea, daeun.choi@kaist.ac.kr; Donghwi Kim, KAIST, Daejeon, Republic of Korea, dhkim09@kaist.ac.kr; Wan Ju Kang, KAIST, Daejeon, Republic of Korea, soarhigh@kaist.ac.kr; Eun Kyoung Choe, University of Maryland, College Park, College Park, Maryland, USA, choe@umd.edu; Sung-Ju Lee, KAIST, Daejeon, Republic of Korea, profsj@kaist.ac.kr.

Proc. ACM Hum.-Comput. Interact., Vol. 5, No. CSCW2, Article 456. Publication date: October 2021.

456

1 INTRODUCTION

As a computer in our hands, a smartphone and its applications (apps) offer a wide array of features. User engagement is an important metric for evaluating products [33], and thus, app providers often take the strategy to ship multiple features in a single "super app" rather than segregating each into a single purpose app [7]. As a result, an app often has a mix of utilitarian and hedonistic features that are tightly intertwined. For example in Facebook, a user may search for an article shared in a group (utilitarian) or watch an entertaining video recommended by Facebook (hedonistic). Addictive features are also embedded in many apps, attracting users into the app and making it harder for users to refrain from smartphone use that they would regret. For example, ephemeral Instagram Stories expire in 24 hours after the post, creating an urge to check the smartphone frequently not to miss any story of their Instagram friends.

Increasing dependence on smartphones raises concerns around overuse out of one's control and unrestrained attachment to smartphones. To promote less regretful smartphone use, intervention designs have been proposed to regulate smartphone use. Restrictive interventions [1, 3, 23–25, 27, 28] lock out the access to smartphone or apps. This approach could be considered impractical due to high dependence on quasiessential features of smartphones and social expectation of constant online connection in relationships and workplace [32]. Non-restrictive mechanisms include providing feedback on one's usage (e.g., the phone- or app-level "screen time" or the number of visits [1, 3, 18, 46]) for self-reflection [18, 40]. However, such feedback is too high-level, not accounting for intra-app usage context. Accordingly, the lack of a nuanced mechanism weakens self-reflection of one's smartphone use and the effect of the intervention mechanism.

Designing an effective intervention against smartphone use that users regret calls for more nuanced analyses of which parts of use users want to gain information about and refrain from, based on the understanding of how phone use experiences vary inside an app. Previous analyses on smartphone use examine various behavior aspects, such as compulsive use [49], habitual use [41], and meaningful use [37]. However, as each work focuses on a specific behavior of phone use, it lacks a comprehensive understanding of which parts of phone or app use the users are satisfied with and which they want to improve.

In understanding various behavioral aspects of smartphone use at a finer granularity, we incorporate the construct of *regret* to explore when and why people feel they "shouldn't have used" their phone. According to regret theory [35], regret is a counterfactual feeling [21] based on a counterfactual inference that "the past might have unfolded differently, particularly if a different decision had been made" [47]. It is aroused when the rewards of a taken action are outweighed by the expected rewards of what could have happened alternatively. Inspecting what people regret presents a retrospective view on which parts of an app are perceived less rewarding and undesirable to repeat and why. As a unit to represent different parts of an app, we dissect phone use by app *features* and analyze regretful use at the feature level, instead of the app or phone level. An app use is broken down into a sequence of *feature* uses; for example, an Instagram session can be split into browsing the feed of following users' posts, browsing the feed of system-recommended posts, direct messaging, viewing stories, and more. We track the use of different app *features* using an app's layout information and prompt users about which cases of feature uses were regretful.

We conducted a one-week experience sampling method (ESM) study with 29 participants using our system, *Finesse*, shown in Figure 1. *Finesse* (i) tracks feature-level uses of four popular social
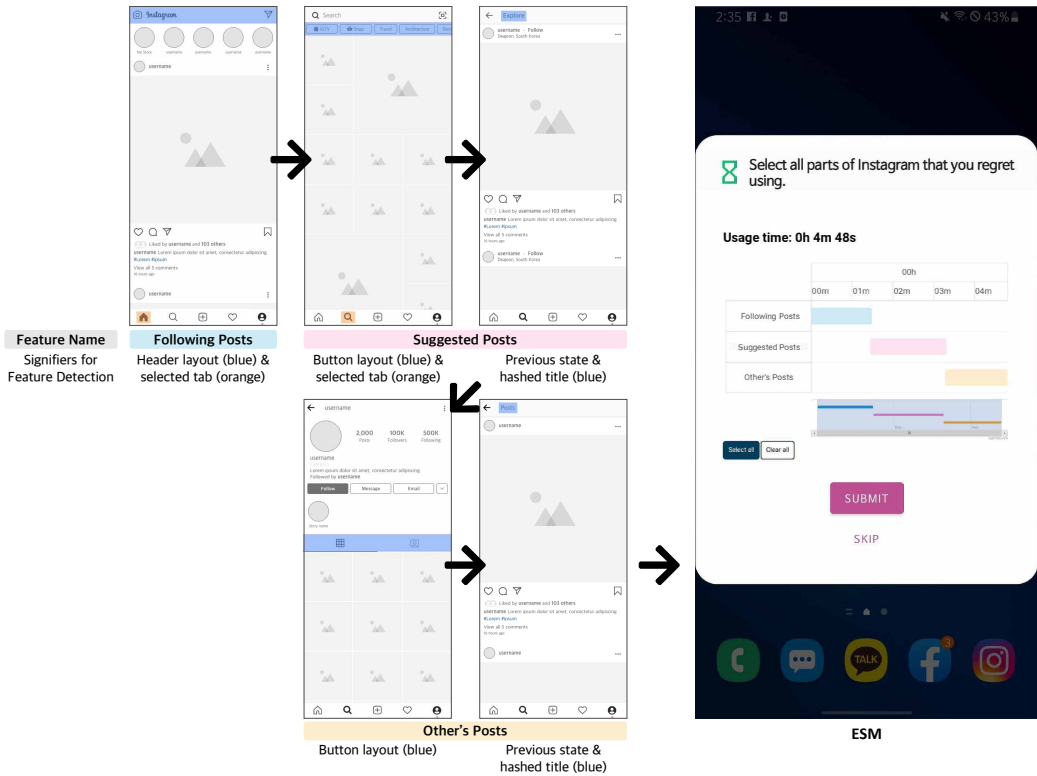
Fig. 1. Example scenario of using Instagram and corresponding ESM screen generated by *Finesse* feature detection. Each screenshot is annotated with the signifiers used to detect each feature state.

media apps (YouTube, Instagram, Facebook, and KakaoTalk) and (ii) prompts users after a session to evaluate which feature uses of the session they regret using. After a week of field study, we performed a retrospective semi-structured interview to learn detailed explanations about the ESM data and collect reflections on cumulative use records. We conducted a mixed-method analysis using the feature-level usage, ESM data, and interview recordings.

Our study reveals different ways that app features are associated with the feeling of regret after smartphone use. First, participants regretted using features associated with habitual checking (e.g., Instagram's FOLLOWING_POSTS, VIEW_STORY, and Facebook's NEWS_FEED) when the features' "following"-based contents are depleted. Second, participants exhibited a habitual feature tour pattern and in-app sidetracking behaviors resulting from the close colocation of instrumental features and attention-grabbing contents recommended by the app provider. Third, participants regretted falling down the rabbit hole following a continuous chain of viewing "just a bit more" of recommended contents. By examining the relationship between app features and regret, we further provide insights into design implications for healthier smartphone use grounded in our findings.

We highlight the paper's contributions as below:

- **Empirical investigations.** This paper reveals feature-level patterns of regretful smartphone usage using regret as the theoretical construct for analysis.

- **Methodologies and tools.** This paper introduces a novel ESM method, *Finesse*, to detect features within an app and display the feature usage log to users for reflection.
- **Design implications.** This paper discusses the implications of our findings on regretful smartphone use and our feature detection-based ESM tool for digital well-being and beyond.

2 BACKGROUND AND RELATED WORK

2.1 Regulating Smartphone Use

Regulatory interventions help users reduce total smartphone usage time and the number of self-interruptions. MyTime [18] is an app that implements three types of interventions: a daily aspiration notification that prompts a user to think of what they want to accomplish that day; a timer that tracks an app usage and displays it in the notification drawer; and a timeout intervention that prompts a "Time's up!" dialog on time limit. MyTime allows users to choose which *apps* to regulate, but not which *features* to regulate within those apps. As a result, it could regulate features that the user desires or fail to regulate ones that they do not.

PomodoLock [23], GoalKeeper [24], and LocknType [25] use restrictive mechanisms for regulating self-interruptions. PomodoLock allows users to set a timer to restrict access to self-interruptive applications and websites. LocknType requires a user to perform a lockout task such as typing a fixed length number before an app use, and GoalKeeper allows users to set a daily use time limit. Lock n' LoL [27] leverages social relationships to provide a phone locking intervention in group activities. NUGU [28] also explored a group-supported self-regulation of limiting smartphone use. Despite the effectiveness of these approaches in discouraging app use, coercive mechanisms triggered frustration in the users due to the lack of nuanced consideration for diverse usage contexts for app use.

Non-restrictive mechanisms based on screen time or the number of visits [1, 3, 18, 46] are phone-level or app-level interventions that are too coarse-grained in today's multi-functional, multi-purpose apps. For example, Instagram serves various features such as browsing the feed of following users, messaging with friends, browsing the feed of system-recommended posts, and searching for accounts or hashtags. We argue that different app features are related to different smartphone use behaviors. An Instagram user might habitually check the feed of following users but use direct messaging with an instrumental intention. Diverse motivations and gratifications of an app use are served through different features, which cannot be captured by a phone-level or app-level analysis.

In web settings, finer-grained interventions exist that restrict or manipulate certain features of a website. Examples include feature minimization tools such as the Facebook news feed eradicator and YouTube recommendations remover [39, 40]. It was reported that a feature-level intervention that restricts only the use of a distracting feature could be effective in helping users stay focused, but an abrupt removal of the feature access might not sustain the intervention effect for a long time due to side effects such as fear of missing out (FOMO). Kovacs *et al.* [30] also underlines the importance of designing interventions based on the understanding of users' mental models to reduce attrition. Therefore, interventions for healthy smartphone use should be grounded in the nuanced understanding of user behaviors and satisfaction in different parts of an app.

2.2 Understanding Smartphone Use Behaviors

A stream of research focused on understanding how people use smartphones, what trigger certain smartphone usage behaviors, and how different behavioral patterns impact user experiences. Oulasvirta *et al.* [41] discovered that people habitually check their phone because they recognize

a quick acquisition of information as rewards. Hiniker *et al.* [19] corroborated that phone use behaviors show distinct patterns between instrumental and ritualistic use of smartphones. Tran *et al.* [49] constructed a descriptive model of the cycle of compulsive checking and identified triggers that start and end compulsive phone use. They discovered that the experience of compulsive phone checking depends on the meaningfulness of the experience, which is mostly determined by informational or social rewards. The notion of meaningful smartphone experience and contributing factors have been analyzed using a Uses and Gratifications (U&G) perspective [37]. They revealed that people have a lower sense of meaningfulness when their phone use was motivated by a habit to pass the time, and the types of use were entertainment and passive social media. Our study focuses on how different functionalities of an app and their design as a *feature* are associated with different behaviors and how they influence the feeling of regret.

A recent study [36] showed how the designs of YouTube's internal mechanisms such as auto-play, recommendations, and advertisements support a sense of agency. It discovered that autoplay and recommendations impair sense of agency while search and playlists support it. It also found that interfaces with greater agency are preferred in an instrumental use. In line with its efforts to inspect into *internal* mechanisms, we present findings on how behaviors of using various UI features attribute to regretful experiences.

Shin and Dey [48] devised an automatic detection of "problematic phone use." They define it as "overuse, undesirable use that results in negative consequences in both personal and social aspects of one's life." Banovic *et al.* [8] categorized smartphone usages into three types based on duration and interaction—glance, review, and engage. Rooksby *et al.* [46] developed a personal tracking tool of screen time and examined the user-side benefits of monitoring their screen time. Hintze *et al.* [20] also tracked the number of interactions and length of sessions in smartphones and tablet devices.

Most of the related work examined the smartphone use in the context of a person interacting with a phone. We broaden the scope of the smartphone use context, incorporating the notion of "regret" to understand what the user could have gained if not using the smartphone, as well as what benefited the user as a result of the smartphone use. Lyngs *et al.* [38] led an interesting discussion about using "regret when reflecting on past activity" as a metric of a user's "true preferences" of behaviors. Although regret is an imperfect metric, using it in behavior analysis provides opportunities to understand what kinds of behaviors users want to avoid and contribute to enhancing self-fulfillment. Nir Eyal proposed the "regret test" [16] as a self-assessment for technologists and designers whether their products are ethically designed without manipulating users and also as regular check-ins about how users feel about their products. We use the concept of regret to investigate perceived rewards of smartphone use and identify the patterns of regretful smartphone use associated with different types of features and rewards.

2.3 Decision Making, Utility, and Regret

To facilitate our analysis on fine-grained behaviors within an app use, we construct our inspection of smartphone use on the concept of *regret*. Kahneman and Tversky defined regret as a "counterfactual emotion" based on a counterfactual inference of what might or should have happened as an alternative to the experienced reality [21]. We leverage this counterfactual, retrospective property of regret to make sense of smartphone use experience and gain insights on how using different features leads to a regretful experience [47].

Behavioral economists have projected the concept of regret to model the behavior of decision making under uncertainty. Decision making theories assume an individual to have a finite number of alternative states of the world, each state j having a probability p_j to occur where $0 < p_j \leq 1$ and $p_1 + ... + p_n = 1$. If an individual makes a choice to take the i-th action among available actions

and moves to a new state j, the individual would have the consequence x_{ij} among an n-tuple of consequences. When making a decision, an individual would consider the *utility* of a consequence, "the psychological experience of pleasure associated with the satisfaction of desire" [35], which would indicate an increment or decrement of wealth in economics applications. Loomes and Sugden [35] posed a regret theory that the choice of an action and its utility is dependent not only on the nature of the chosen action's consequence but also on the nature of other alternative consequences. For example, when an individual experiences a consequence x_{1j} following an action, whereas the consequence of another action x_{2j} turns out to be more desirable, the individual may experience regret, leading to a reduction in the psychological utility of the consequence x_{1j}. In other words, if the best predicted rewards among all alternatives exceed the actual experienced rewards after a choice, the choice becomes regretful.

We leverage regret theory to understand smartphone use at the app feature level. As a theory rooted in traditional behavioral studies, regret theory serves as an established construct with concrete factors that attribute to the feeling of regret. Since regret is a cognitive emotion resulting from comparing the rewards of smartphone use and alternative choices, analyzing regretful experiences facilitates knowing the perceived rewards of smartphone use and when and why the rewards become insufficient to leave no regret. Based on our experience sampling and interview data about regretful smartphone use, we explain behaviors related to different features, perceived gratifications of different feature uses, and different pathways to regret in different features. '

3 METHOD

We aim to examine (1) immediate regret after smartphone use, (2) types of rewards gained from social media app use and alternatives rewards to smartphone use, and (3) detailed accounts of regretful uses and reflection on accumulated use. The immediate regret was measured via the experience sampling method (ESM) over 7-day period, and we relied on the retrospective interviews to draw immediate and alternative rewards of smartphone use and rationale behind regretful smartphone uses. We developed *Finesse* to collect feature-level use data and sample regretful experience *in-situ* of popular social media app use. We chose Instagram, Facebook, YouTube, and KakaoTalk,[1] which are four of the top five apps in the time spent monthly on average in Korea [11]. The selected apps are often associated with problematic use [14] and contain multiple features with varying motivations of use, ranging from purely hedonistic purpose to productive. In the following, we first summarize the participant demographics and target app selection. We then explain how we designed and implemented *Finesse* to assist the ESM study followed by details of the ESM study method and the retrospective interview. The user study was approved by the Institutional Review Board (IRB) of the authors' institution.

3.1 Participants

Through university online forums, we recruited 29 Android users (female = 15, male = 14; mean age = 21.97, min age = 19, max age = 27) who met our inclusion criteria of using at least three out of the four target apps on a daily basis. We targeted the 18-29 age group as they are the most active users of social media [4] and more likely to report poor mental health [14]. We required the daily usage of at least three apps to capture smartphone use experiences that could vary by apps and features even for the same user, depending on the user's purpose and patterns of usage. The group consists of 21 undergraduates, seven graduate students, and one unemployed. Among the 29

---

[1]KakaoTalk [13] is a dominant mobile messenger app in Korea [22]. Besides the primary chat feature, it serves other features such as news, shopping, bill payments, etc.

participants, four do not use Facebook and two use it fewer than once a month. All participants use the other three applications more than once a day.

3.2 Finesse: Feature Use Tracker

Finesse tracks the feature-level use of YouTube, Facebook, Instagram, and KakaoTalk. We describe the criteria used to break down each app into features, how we implemented the feature tracker to work on commodity apps on commodity phones, and the design issues that are taken into account.

3.2.1 Criteria for Features. We extracted the features of four target applications by considering the combination of two criteria: (1) user actions and (2) nature of the contents. First, we grouped the features based on user action such as viewing, searching, uploading, and chatting. The nature of the content includes the *form* of contents (e.g., videos, feed, notification, and posts) as well as the *source* of the content (e.g., user's own posts, posts from whom the user follows, and posts from others the user does not follow). Considering these two criteria, we defined 43 features from the four applications. For example, we defined Instagram's SUGGESTED_POSTS as a feature for *viewing* (action) the *feed of posts* (form of contents) with *system-recommended* (source of contents) contents. All features are listed and labeled with the criteria in Appendix A.

3.2.2 Design and Implementation. We used the Android Accessibility API [15] to detect the feature currently in use. We retrieved tree-structured UI components information for every scroll, click, and focus event. We selected some of this information for feature use detection. The selected element in the menu bar is a good indicator of the feature in use. We also used layout information of each component. We used class names of activities and components given by developers to detect more complex features. The window ID and previously detected features are used to analyze the use of various features in a session. Lastly, we used hashed text data, such as contents description, for minimal use. We could not possibly cover the exhaustive list of features in the four target applications. Therefore, if a new layout other than the pre-determined features was detected, it was labelled as undefined.

3.2.3 Privacy. As *Finesse* monitors all screen information to detect feature use, we must consider privacy issues. The data that *Finesse* collects do not contain information about the content on the screen or information that reveals the subject, except the text information. We used two types of text information—text from contents and content description (added by developers about the purpose of UI elements). To minimize privacy risk, we used all text information in a hashed form. We used this information only to understand if certain words or phrases were present within the screen. For example, we used the hashed text to check whether the word *post* is present in the header to determine whether the current feature is related to viewing posts.

3.2.4 Compatibility. We built *Finesse* to be compatible with most Android devices. We configured the application to work consistently regardless of various screen sizes. In detecting feature uses, we refrained from utilizing UI component size information as it could vary depending on the device. Instead, we mainly used the class name of the component as these are identical across all devices. *Finesse* detects feature uses by catching the pre-determined indicators from the screen. However, as we cannot predefine an indicator for every combination of on-screen contents, unassigned cases are labelled as undefined features. We note that, while using some predefined feature A, an undefined feature U might be detected and wrongly inform *Finesse* that the pre-defined feature A has terminated. To prevent this unexpected discontinuity, we used a hash map that matches the window ID with the detected feature A so that if the user views the screen with the same window ID, *Finesse* considers that the same feature A to be still in use.

3.3 Experience Sampling

Finesse runs as a background service and collects phone use logs and ESM response per app session. An app session starts when a target app is opened. The session ends when the user transitions to the home screen, the foreground app switches to another via push notification, or the phone is locked. For every session, it stores the following attributes in our cloud server: the session start time, session duration, and list of feature instances (feature name, feature start time, and feature end time for each instance) used in the session.

Leveraging the tracked feature uses, we collected *in-situ* samples of regretful phone use experiences through a prompt right after the end of a target app session, asking a participant to select instances of regretful feature use, as shown in Figure 1. Because regret is a retrospective feeling, asking whether one regrets right after app usage facilitates evaluation of the experience while minimizing interference in the middle of app use experience.

We aimed to minimize the effort to answer our ESM to prevent the ESM prompt from acting as an intervention to encourage self-reflection. Asking for a user response every session would be ideal for data collection, but it would incur excessive user burden and likely result in sloppy responses. Purely random sampling might result in a skewed sample set towards short, frequent uses in a certain period of time, which is likely to represent habitual checking behavior and short chat message replies. Therefore, we devised a sampling policy to balance the four target apps, session duration, and use duration across samples. We divided session duration into three ranges: duration < 30 seconds, 30 seconds ≤ duration < 5 minutes, and 5 minutes ≤ duration. The 30-second divider classifies short duration [41], and the 5-minute divider represents the average app session length of social networking apps [12]. To balance the time of use in a day, we also divided a day into eight 3-hour bins (*e.g.*, 0 to 3AM, 3 to 6AM, and so on). Our sampling prompt appears at most once per session duration range, per target app, in each 3-hour bin with a random 50% probability. We skipped sampling sessions shorter than 5 seconds to prevent samples of mistakenly opened sessions.

For the sampled sessions, *Finesse* displays an overlay dialogue (Figure 1, right) immediately after the end of the session. The prompt visualizes the feature uses of the session in a timeline graph where the horizontal axis marks the time from the session start to the end; and each row in the vertical axis corresponds to the features used in that session. The maximum range shown in a window is 5 minutes, and participants can scroll from the beginning to the end using the navigation bar at the bottom of the timeline in a 5-minute window. The prompt asks the participants to select all feature instances they regret using by clicking the item(s) and submit their response. If the prompt is not answered within 5 minutes, it automatically dismisses so that the ESM disregards non-immediate regret and captures only immediate regret after use.

Prior to the 7-day experience sampling period, all participants attended a 1-hour app installation and tutorial session via Zoom. We grouped a maximum of five participants into sessions based on their availability and smartphone models to facilitate the instruction process. At the beginning of each session, the experimenter explained what types of data will be collected in the study: all screen layout information, encrypted text, feature-level usage time of the four target apps, ESM responses, and personal information for payment. We informed participants of their right to withdraw at any point. Only after all participants in the session typed "I agree." in the chat, the experimenter continued the instructions. After the session, we asked all participants to carefully read the IRB consent form and return the signed copy if they agree to participate. No participant dropped out for a data collection-related concern.

3.4 Retrospective Interview

After one week of experience sampling, we conducted a retrospective semi-structured interview to investigate the reasons behind regret and collect reflections on aggregated data of feature uses. All interviews were conducted individually and remotely via Zoom for an hour. Questions asked during the retrospective interview include: Why did you report a certain instance of feature usage as regretful? Not all usage instances of this feature are regretful; what's the difference? How would you have used it differently to not regret? What makes you use the same feature again despite reporting as regretful? To help participants recall the context of an ESM response, we created a dashboard that visualizes (1) a collection of use cases that the participant answered as regretful and (2) the weekly cumulative use. The collection of regretful cases include: the count and percentage of regretful use per app, the count and percentage of regretful use per feature, and the duration and count of regretful use in hours of a day and days of the week. The weekly cumulative use contains the total duration of each feature per app, the total count of each feature per app, and the duration and count of use per app and per feature in hours of a day and days of the week.

3.5 Data Analysis

We performed a mixed-methods analysis on phone use logs, ESM responses, and interview data. For quantitative analysis, we used phone use logs and ESM responses to calculate the feature use probability for each app, feature regret ratio for each app, and feature use count in each in-session time bin. The time bin analysis illustrates the in-app user trajectory that could better elucidate factors relevant to regretful smartphone use experience. It conjectures that, even for two sessions with the same duration, their in-session share of feature uses may be of drastically different composition.

For qualitative analysis, we followed the procedure of thematic analysis [9] and applied the constant comparative method [17]. We first transcribed the interviews conducted in Korean. Four authors individually coded two contrasting samples of the interviews and then discussed them together to share initial codes and potential themes, which were later turned into a codebook. The lead author coded the remaining interview data while continuously refining the codes and updating the codebook. Upon completion of the first round of coding, another author coded one interview sample based on the updated codebook for verification. Throughout this process, all authors met regularly to discuss potential themes while triangulating the interview data with the quantitative analyses. Bilingual authors translated the quoted statements used in the paper.

4 RESULTS

We present our results analysis on regretful smartphone use. We first provide a summary of collected app use sessions and ESM data. We collected a total of 18,263 app use sessions (daily average of 90.0 sessions per person) of the four apps in 7-day data collection window from each 29 participants. Among these, 4,246 sessions were sampled for ESM and participants answered 4,069 of them. ESM response rate was above 92% for all days with an overall average of 95.8%. Of those responses, 26.5% of sessions (1,079 sessions) were considered fully or partially regretful as they contain one or more regretted features. Of the 1,079 sessions, 431 (39.9%) were fully regretful, meaning that all feature instances of the session were selected as regretful, and 648 sessions (60.0%) were partially regretful. Table 1 summarizes the total number of all sessions, daily average of sessions per person, number of the answered sessions, and number of sessions answered as regretful per app.

Figure 2 shows the feature use probabilities for each feature of an app. It plots each individual's feature use probability in a box plot. The entry features (Instagram's FOLLOWING_POSTS, Facebook's NEWS_FEED, YouTube's BROWSE_HOME, and KakaoTalk's CHAT), which first appear upon entrance

App	Total Sessions	Daily Average Per Person (σ, M)	Answered / Total ESM	Regretful
Instagram	3,540	17.4 (19.5, 12)	945 / 985	396 (41.9%)
Facebook	1,880	10.7 (11.1, 8)	603 / 622	200 (33.2%)
YouTube	2,366	11.7 (15.8, 7)	663 / 693	146 (22.0%)
KakaoTalk	10,477	51.6 (31.1, 48)	1,858 / 1,946	337 (18.1%)
Total	18,263	90.0 (49.2, 82)	4,069 / 4,246	1,079 (26.5%)

Table 1. The number of app sessions collected during a one-week study from 29 participants. Total Sessions include ones with or without ESM questionnaires. Daily Average Per Person is the daily mean of all sessions with or without ESM questionnaires per person along with standard deviation (σ) and median (M) values in parentheses. The fourth column is the number of answered sessions out of total sessions with ESM. The last column (Regretful) is the number of sessions answered as fully or partially regretful and its percentage out of all answered ESM sessions in parentheses.

to an app, naturally have the highest probability of use near 1.0. It is not a 100% because there are cases where a session resumes from a feature that the user left off in the previous session. Individual differences are notable for several features; for example, Instagram's SUGGESTED_POSTS has a variance from 0.0 to approximately 0.75. It suggests that while some participants barely look at the SUGGESTED_POSTS in a session, others browse the feed three out of four times. This feature use probability can be used as a reference to interpret future findings.

We next present our findings on (1) the types and dynamics of reward associated with different features, (2) features that account for regretful use and reasons behind it, and (3) patterns of regretful use based on the feature-level ESM samples and interview data.

4.1 Decomposing Regretful Use

As presented in Section 2.3, whether an action concludes in regret is a function of utility, which in regret theory is associated with the consequence of the taken action (actual reward denoted as x_A) in comparison with the best anticipated consequence of alternatives (best reward denoted as x_B). The key factor of regret is thus how *desirable* the actual reward x_A is in comparison with the best reward x_B. If x_B is a more desirable consequence than x_A, an individual might experience regret. The feeling of regret arises as a combination of how desirable the rewards of social media app use were and how desirable the rewards of alternative actions might have been. Since we sampled the immediate regret after each app use session, in the context of our study, the actual rewards correspond to the consequence of app use, and the alternative rewards correspond to other tasks that the participant would have done, had they not used the app. We summarize different reward types of social media app use and of alternatives, which we captured during the retrospective interview.

4.1.1 Rewards of Social Media Use. During the retrospective interview, participants described cases when they did *not* regret using social media apps due to various rewards they received. We inductively identified four types of rewards that participants reported to be desirable. The four types of rewards are not necessarily mutually exclusive.

Social rewards refer to the value of staying connected to friends through direct messaging or catching up on their lives. For example, active sharing and keeping records of daily life through Instagram bring social rewards by "building memories and sparking direct communication with friends," as P2 mentioned.
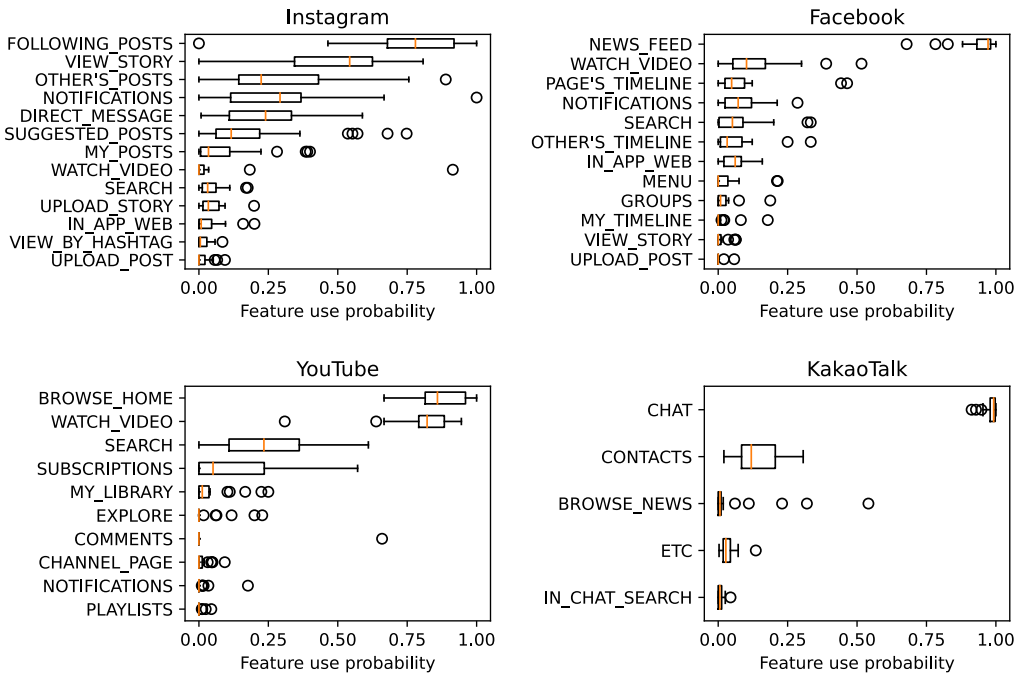
Fig. 2. A box plot of feature use probability for each app. The x-axis is the feature use probability (0-1). The y-axis represents each feature name of the app. The orange line in the box plot represents the median number.

Informational rewards are the acquisition of new information from social media use. Users obtain new information through a search about a specific topic. Users could also opportunistically learn new information through informative contents. They could learn new information ranging from recent news stories about COVID-19 to trending Netflix movies.

Personal interests refer to the feeling of satisfaction gained from the social media contents that correspond to one's interests. For example, P13 described how Instagram yields rewards for personal growth: *"I like dancing, so videos of dancing are usually listed [in the* SUGGESTED_POSTS*]. If I click [a user's account that leads to* OTHER'S_POSTS*] to watch that person's dance, it's beneficial so I didn't check [that I regret]. I also didn't regret viewing other's page related to fashion or studying."*

Entertainment rewards refer to hedonic rewards that bring pleasure through fun and entertaining content. To paraphrase P8 as an example, he intentionally and consciously watched videos such as entertainment shows for fun to refresh his mood.

In Sections 4.2 and 4.3, we elaborate on how these rewards change throughout smartphone use and how these dynamics result in the feeling of regret in various pathways.

4.1.2 Alternative Rewards to Smartphone Use. How much rewards one gained from smartphone use is not the sole factor influencing the feeling of regret. The regret theory suggests that, when looking back on the past, if possible rewards of alternative actions exceed the rewards of social media use, an individual would regret the use. Accordingly, participants frequently reported in the interview that whether they regret it depended on the primary task at the time of use. The presented findings are summarized from their responses to our interview questions "Why did you

report a certain instance of feature usage as regretful?", "How would you have used it differently to not regret?", and follow-up questions.

Possible alternative rewards include better productivity, physical conditions, and social relationships. As most participants are students, using social media apps often led to breaking concentration while they were "studying", "doing assignments", or "taking lectures". Had they not used the phone, they would have maintained focus and been rewarded with increased productivity. Participants often regretted the use before sleeping and during a social meal. Had they not used the phone, they would have fallen asleep earlier, slept longer, and had better physical conditions the next day. For example, when the smartphone use "affected schedules or tasks to do the next day," P24 regretted "that 'I should've woken up early tomorrow' or that 'I can only sleep for five and a half hours if I sleep now.'" During the day, P11 would have "rather taken a nap for rest." During a meal, participants would have focused on conversations, "not violated the social norm of using the phone during a meal" (P26), and had a more positive impact on the social relationships.

In Section 4.3, we explain how the potential rewards of alternative actions to smartphone use play a role in regretful smartphone use.

4.2 Features and Regret

Figure 3 shows a box plot of each feature's individual regret ratio in descending order, where a feature with the highest average regret ratio appears at the top. A feature's regret ratio is the count of feature use instances selected as regretful, divided by the total number of feature use instances in the ESM questionnaires. The box plot illustrates individual differences in the perception of regret for each feature.

Figure 4 and Table 2 (in Appendix C) summarizes the results of the general linear mixed effects regression analysis to extract feature-regret relationships in each app. The analysis sets each feature as a fixed effect and participant ID as a random effect. It runs `glmer` models with the maximum likelihood estimation and significance test using R and the lmerTest package [31]. For each app, we used the app's primary feature that first appears on entry and was most likely to be used across all ESM sessions (Figure 2) as the reference baseline factor to compare the primary feature and other features of the same app for their effects on regret.

The results in Figure 3 and Figure 4 (and Table 2) show that features in the same app have varying associations with regret. Overall, it is observed that features that comprise *active* form of usage (labeled as chatting, searching, and uploading in User Action of Appendix A, e.g., Instagram's `UPLOAD_POST`, `UPLOAD_STORY`, `DIRECT_MESSAGE`, KakaoTalk's `CHAT`, and `SEARCH` of Instagram, Facebook, and YouTube) are less associated with regretful experiences than features related to *passive* usage patterns (labeled as viewing in User Action of Appendix A, e.g., Instagram's `FOLLOWING_POSTS`, `SUGGESTED_POSTS`, `VIEW_STORY`, Facebook's `WATCH_VIDEO`, `NEWS_FEED`, YouTube's `BROWSE_HOME`, `EXPLORE`, KakaoTalk's `NEWS`, and `CONTACTS`). P28 reported less regret when having a clear objective of seeking information: *"When I use `SEARCH`, I have an objective to find this information, so I regret less than just watching the `NEWS_FEED`."*

Note that features serving the same types of rewards vary in their effects on regret when the ways of presenting rewards are different. Taking Instagram as an example, `VIEW_STORY` and `FOLLOWING_POSTS` both offer social rewards with similar types of contents—stories and posts from a user's following accounts. However, `VIEW_STORY` has a higher regret ratio than `FOLLOWING_POSTS`. Because a story is ephemeral and *"disappears after 24 hours,"* if there's a circle that says they haven't seen it yet, it triggers an urge to *"just press it"* before it is gone. Some participants displayed a regretful act of fidgeting with the `VIEW_STORY` feature. P14 explained: *"I keep pressing next and flipping a story to another. I just keep pressing... to just waste time rather than actually viewing it. I regret that I'm doing this again after a while."* This example illustrates how different interaction
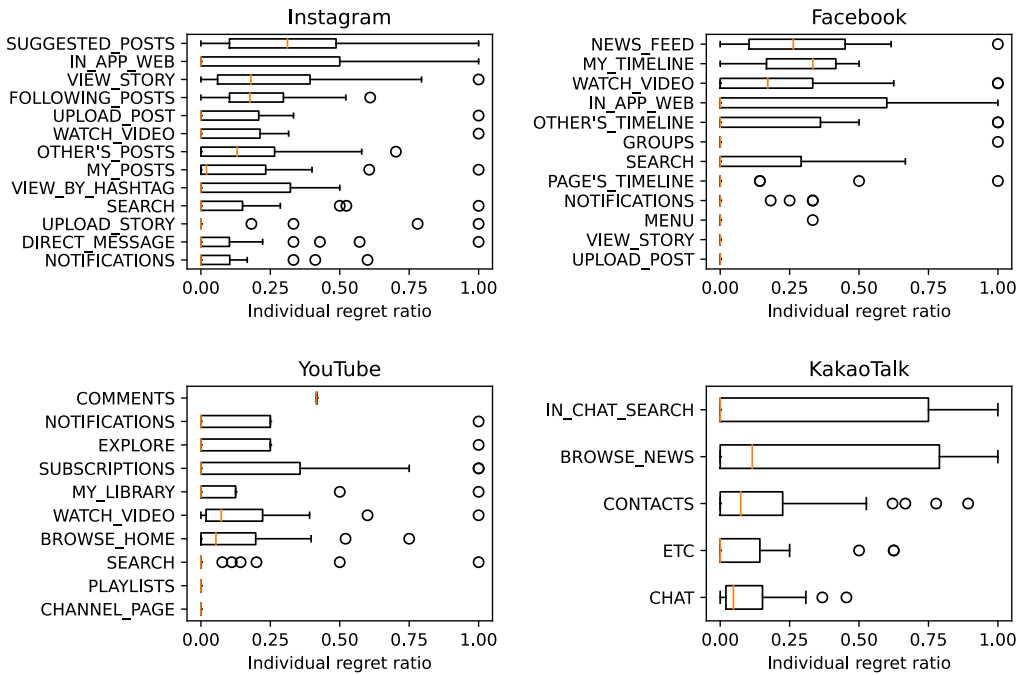
Fig. 3. A box plot of feature regret ratio for each app. Each point in the box plot refers to each participant's feature regret ratio for each app. A feature regret ratio is calculated as the count of feature uses answered as regretful divided by the total number of feature uses that appeared in the ESM. The x-axis is the feature regret ratio (0-1). The y-axis represents each feature name of the app. The orange line in the box plot represents the median.

dynamics of features influence one's regretful usage behavior even when the contents and expected rewards are similar.

Features with similar ways of presenting rewards but different types of rewards also showed significant difference in the effects of regret. Instagram's FOLLOWING_POSTS and SUGGESTED_POSTS both employ an infinite scroll-based feed form to present posts of users. However, they have different sources of posts: FOLLOWING_POSTS shows posts of the users that one is following, thereby usually containing information that a user has expressed preferences to subscribe, whereas SUGGESTED_POSTS shows posts that are recommended by Instagram's algorithms, thereby containing *"random"* or *"meaningless"* posts. These examples highlight that both the expected rewards and the dynamics of reward from using a feature affect a user's behaviors and feeling of regret after use.

We next provide detailed explanations on why and how features differ in their regret ratios in relation to the types of rewards and the dynamics of reward presentation to users.

4.3 Feature-Level Patterns of Regretful Use

Figure 5 visualizes what features participants used at what time in a session for regretful (Figure 5 (a)) and non-regretful uses (Figure 5 (b)). Time within a session is divided into six time segments (0-10s, 10-30s, 30s-1m, 1m-5m, 5m-10m, and 10m-) based on the distribution of samples, and each feature use is counted in each time segment for all sessions. We use the count fraction instead of time because participants selected which feature segment they regret rather than specifying the exact

(a) Instagram (Baseline: FOLLOWING_POSTS)

(b) Facebook (Baseline: NEWS_FEED)

(c) YouTube (Baseline: BROWSE_HOME)
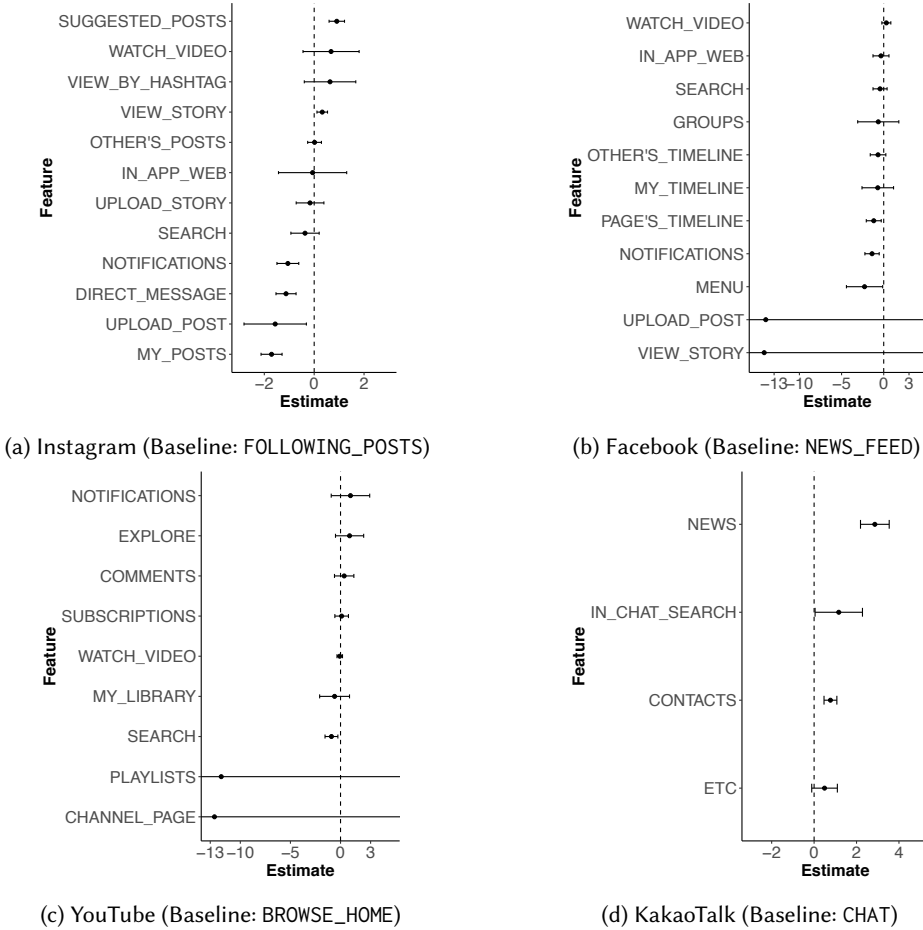
(d) KakaoTalk (Baseline: CHAT)

Fig. 4. Caterpillar plots of general linear mixed model (GLMM) analysis results on features and regret. Refer to Table 2 in Appendix C for the numerical coefficients and more values. Each app's caterpillar plot displays the feature coefficients and their 95% confidence intervals in comparison to the app's baseline feature. Features that are right of the center line are *Estimate* times more likely to be regretted after use in comparison to the baseline; features left of the center line are *Estimate* times less likely to be regretted than the baseline feature. For example, in the case of KakaoTalk, relative to CHAT, participants were 2.83× more likely to regret using NEWS. Note that the baseline features are each app's primary feature that first appears on entry and was most likely to be used across all ESM sessions.

time frame in the ESM prompt. There are fewer samples for longer sessions such as >5m and >10m, and the lengths of longer sessions are more sparsely distributed along time. Thus, usage after 10 minutes are aggregated to avoid overfitting to a small set of samples. From the participants' explanations on when and why they regretted, we extracted three patterns of feature-level behaviors producing regretful use. We refer to this figure to explain each pattern of regretful use and describe how it is exhibited in the actual usage data.

4.3.1 Habitual checking quickly depletes rewards from feature use. Habitual checking is a habit of opening social media apps and checking feed-based or story-based features on entrance, such as

(a) Regretful uses only.

(b) Non-regretful uses only.
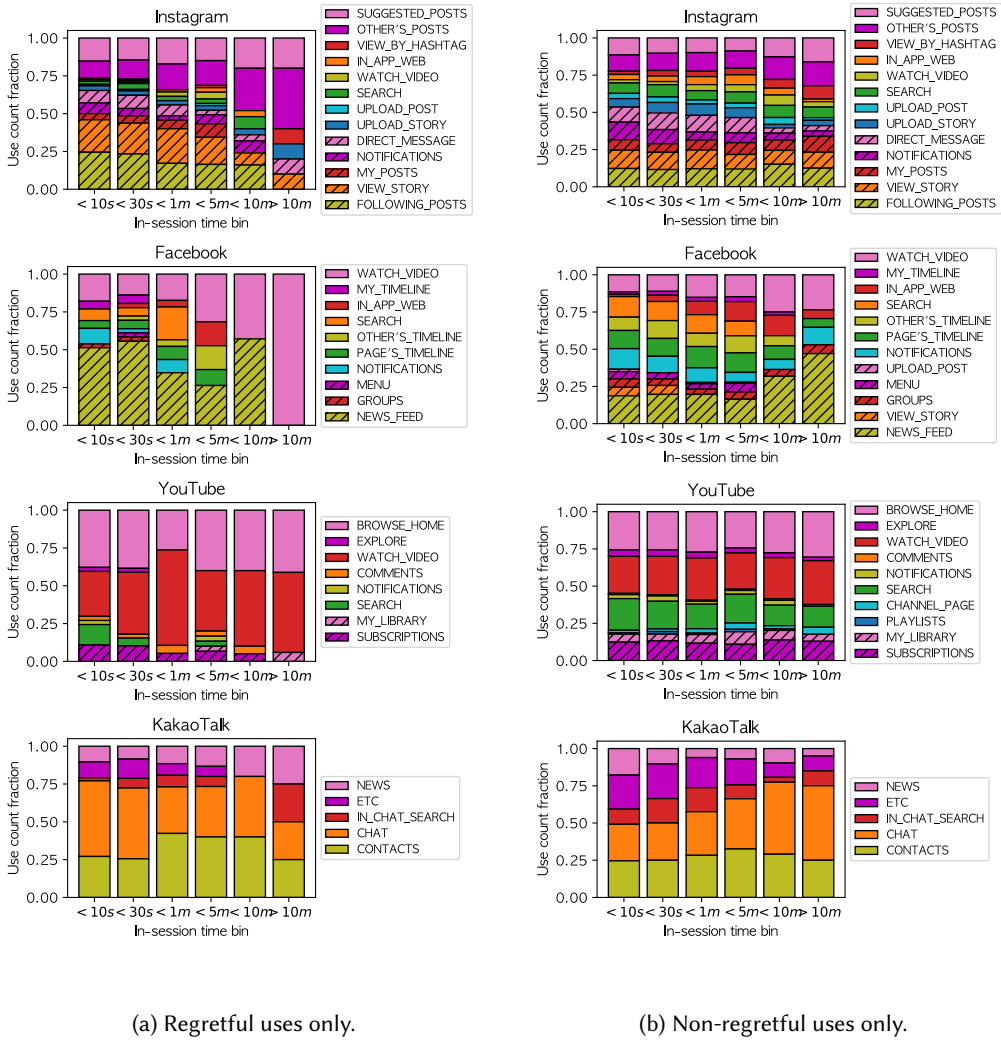
Fig. 5. Each app feature's use count share for each in-session time bin. Subscription-based features such as Instagram's FOLLOWING_POSTS, VIEW_STORY, Facebook's NEWS_FEED, GROUPS, and YouTube's SUBSCRIPTIONS are grouped at the bottom of the stacked bars. Recommendation-based features such as Instagram's SUGGESTED_POSTS, Facebook's WATCH_VIDEO, YouTube's BROWSE_HOME, and KakaoTalk's NEWS are grouped at the top. The groupings of subscription- and recommendation-based features are available in Appendix A.

YouTube's BROWSE_HOME, Facebook's NEWS_FEED, Instagram's FOLLOWING_POSTS and VIEW_STORY. In line with the previous finding [37] that habitual motivation of smartphone use is associated with lower meaning, many participants reported that the habitual checking pattern is one of the main reasons of regret. However, habitual checking does not *always* leave a user with regret. P21 explained not regret using Facebook's NEWS_FEED when it provides informational rewards: *"I*

entered [Facebook's NEWS_FEED*] out of habit, but I don't really regret when I see several news and learn new facts."*

Participants felt regretful for habitual checking when it occurs repeatedly and *fails to deliver new content*, hence no reward. P4 explained: *"I think it was because there wasn't anything new when I just scrolled it [Facebook's* NEWS_FEED*] down when I was bored. I was often on and off of it again and again, so the interval was short. I saw it about 5 minutes ago, but I was on it again. I regret having watched it again without new contents."*

Habitual checking with little or no rewards is sometimes followed by a disoriented reward seeking behavior where users linger on the feature, mindlessly seeking for some rewards. P14 stated *"I don't subscribe to many channels, so not many videos are there [in YouTube's* SUBSCRIPTIONS *feature tab]. While I am browsing [the list of videos by subscribed channels], if I don't find anything new, I should get out and watch other videos. But I just keep spending time there even though I've seen those thumbnails multiple times. If I don't find a video I want to watch in the* SUBSCRIPTIONS, *I feel regretful."*

This pattern of rewards depletion accompanying regret mainly appears with the use of *subscription-based* features (labeled as 'subscription-based' in the Source of Content column of Appendix A). These are the features that deliver contents uploaded by users, pages, or channels that a user is following or subscribing. Examples are Instagram's FOLLOWING_POSTS, VIEW_STORY, Facebook's NEWS_FEED, and YouTube's SUBSCRIPTIONS. Because the content providers for subscription-based features are other users or pages that the user follows, there could be a finite amount of content available at a time. Repeated uses thus quickly deplete contents and cause regret.

This pattern is observed in Figure 5. For the ease of interpretation, we grouped subscription-based features at the bottom. In 'Regretful uses only' (Figure 5 (a)), subscription-based features (Instagram's FOLLOWING_POSTS, VIEW_STORY, Facebook's NEWS_FEED, GROUPS, and YouTube's SUBSCRIPTIONS) all take a larger portion of regretful uses towards the beginning of a session. However, in 'Non-regretful uses' (Figure 5 (b)), their use count fractions remain consistent throughout a session. This corresponds to the users' explanations of regretting when they encounter lack of contents and rewards on entrance and lingering.

4.3.2 Deviation from original purpose of use. When participants failed to earn satisfactory rewards from habitual checking and lingering, they imparted cases of a 'feature tour' wandering around other features within the same app. P22 habitually turns the direction to other features in Instagram when expected rewards are not satisfied in Instagram's FOLLOWING_POSTS: *"I just habitually swipe to [Instagram's]* SUGGESTED_POSTS *if there's no [posts in]* FOLLOWING_POSTS. *It's more diverse, so I browse it for longer."*

The pattern of a habitual feature tour also arises when a user starts an app session with instrumental motivation. For example, users usually have a strong primary motivation for using the CHAT feature of KakaoTalk. After completing the original purpose of using the app, however, several participants execute their unconscious habit of hopping on to other features, which often results in regret. P14 described a routinization of this feature tour:

"When I open KakaoTalk to read messages, I check the news most of the time. Or sometimes NEWS *appears up front when I didn't close my app properly the last time. I can't help but use [KakaoTalk] because there are a lot of chat rooms. After chatting, I tend to swipe over unconsciously as if it's a must-do thing. [...] For* ETC, *there's really nothing to do. I just check how much money I have. I look at some gifts options and then come out. It's nothing meaningful, so I regret it a lot."*

KakaoTalk's NEWS shows an infinite stream of recommended news articles, and ETC comprises of diverse mini features such as online payment, online gifts shopping, etc.

Such sidetracking from the original intention of use also takes place in *prior* to using the intended feature due to distraction. P5 shared an example: *"I didn't open Instagram to see the* VIEW_STORY*, but to read the feed [*FOLLOWING_POSTS *or* SUGGESTED_POSTS*] or send a* DIRECT_MESSAGE*. I habitually tapped on* VIEW_STORY*, and I regretted it because I wasted time."*

As a result of deviation from the originally intended use, participants often landed on features with infinitely available posts or videos. As opposed to subscription-based features, these system recommendation-based features (labeled as "recommendation-based" in the Source of Content column of Appendix A) serve more random and unpredictable rewards. According to regret theory, sidetracking from the originally intended use to recommendation-based features is likely to result in regret because the expected, actual rewards of their intended use are delayed while unintentionally elongated use increases the cost of time and thus alternative rewards.

The reported cases of deviation from original intention were often unconscious or hard to resist due to the app's workflow design in which features of active/passive usage or instrumental/ritualistic usage are mixed in between. Instagram, for example, places the active features (e.g., SEARCH), which are often intentional and desirable, and recommended passive features (e.g., SUGGESTED_POSTS), which are less intentional or desirable but tempting, on the same page. P3 reported:

> *"I can see it [*SUGGESTED_POSTS*] when I go into that magnifying glass [button]. I think it's pretty tactical because it leads someone to follow through a catchy post before they actually search. When I go there to* SEARCH*, I'm curious about this [post on the feed], but it would disappear when I come back from searching. So I just click once to see it, but I don't go in voluntarily; it's not like 'I should go check the* SUGGESTED_POSTS*.' "*

While the app designers are successful in engaging users for longer duration, this results in users falling into distractions and having a regretful experience.

4.3.3 "Just a bit more" brings users down the rabbit hole. The act of deviation from the original purpose of use, described in Section 4.3.2, sometimes induces a prolonged use, bringing users down the 'rabbit hole'. P1 explained how a habitual feature tour is followed by a prolonged use: *"When I go in [to Instagram], I go through the* SUGGESTED_POSTS *at least once. When I press it, if I see something I'm interested in, I use it longer. Otherwise, I quit the app right away. I have a habit of checking it automatically when I go in"*. In this example, the rewards of unintentionally prolonged app use (using SUGGESTED_POSTS) are low, whereas the originally anticipated rewards are delayed. Therefore, participants felt regretful about the prolonged use induced by a deviating use.

Participants noted that it is especially hard to break the chain of viewing recommended contents although they wanted to keep the usage short. P24 pointed out that because contents are bite-sized, it is hard to resist the thought of using *"just a bit more"*:

> *"I always start off thinking I should watch just one short video before bed, but it doesn't really turn out that way. Maybe I shouldn't start watching videos at all before bed, but it's easier said than done. [...] YouTube videos aren't that long these days. A lot of them are 5 to 10 minutes long, and since it's short, I often think that I can probably watch just one more. Of course, it would be nice if one remains strong-willed to watch only one video and go to sleep. But when I didn't have enough free time for myself during the day, I think it would be okay to watch just one video at night, and then I end up staying up late."*

P29 also denoted the infinite stream of videos that are played out without effort: *"While viewing* NEWS_FEED *on Facebook, when I scroll down, videos keep popping up. I regretted watching the videos [in* WATCH_VIDEO*] and losing track of time."*

As exhibited in these examples, "just a bit more" behaviors push users down the rabbit hole of endless feed with bite-sized items. Falling into the rabbit hole delays what they intend to do after smartphone use and results in regret. This pattern is conspicuous in Figure 5. To facilitate interpretation, the recommendation-based features are placed towards the top of the stacked bars. Features regretted near the end of long sessions (>10m) are dominated by features such as YouTube's BROWSE_HOME, WATCH_VIDEO, Facebook's WATCH_VIDEO, KakaoTalk's NEWS, and Instagram's SUGGESTED_POSTS and OTHER'S_POSTS (often a result of profile hopping from SUGGESTED_POSTS). These features feed an endless stream of recommended posts and videos.

4.4 Effects of Reflecting on Feature-Level Usage

In our retrospective interviews, we showed each participant a feature-level usage report for their weekly use, visualizing the total time spent on each feature and the total count of each feature segment by app. Participants also shared their experiences of reflecting on the feature-level summary of a session usage in our ESM (Figure 1, right). Participants, especially those who utilized various features of an app, appreciated the feature-level usage information over app-level information for three reasons.

First, feature-level usage information helps identify how much time was spent out of necessity and how much was not, especially when "the purpose of usage is extremely different such as KakaoTalk's CHAT and NEWS" (P27). P3 explained:

"I didn't know which features I use a lot and how much time I spend on those. There was no way to get such information. I think there are parts in an app that everyone knows are meaningless to use. In Instagram, especially the SUGGESTED_POSTS, *it really doesn't help your life at all, and I know it. It's hard to track exactly how much I use it and when I use it. But if I look at it [feature-level usage report], I certainly think 'Ah, I indeed use this feature a lot.' Such feedback can be useful in motivating me (to change behavior)."*

Second, viewing feature-level usage information prevents self-justification on usage with evidences of unnecessary use. It makes users be "self-aware of usage with unnecessarily large portion" (P6) while they would "pay less attention" to an app-level usage report which aggregates both necessary and unnecessary uses. According to P5, *"it [the app-level screen time] includes things like messaging, so I'll probably justify myself that it's okay because I was looking at the messages. But if it says, "You viewed the news feed for 10 minutes", I'll feel "I didn't even mean to see this but spent 10 minutes." This way I'll be self-aware and want to control it."*

Third, seeing feature-level usage information helps design an actionable plan on which part of app usage they should strive to reduce. P1 reported:

"I think seeing [the screen time] by features certainly provides specific evidence for determining whether you did it because you needed. For example, in KakaoTalk, I don't feel bad about using [the time] for conversation. But if I spent about an hour a week checking people's profiles [in CONTACTS*], it makes me think I should change those behaviors. So seeing by features will help me a lot more to think of a solution to change my life pattern. Simply saying I used it for several hours is not specific."*

Showing such personal feature-level summary of usage could induce the feeling of regret in a short run but helps users "better reflect" (P26) based on "numerical quantification of slacking off" (P3) and make an actionable plan for reduction in the long-term regret.

4.5 Summary of Key Findings

- Four main rewards of social media app usage are social, informational, personal interests, and entertainment rewards; and alternative rewards to smartphone use include productivity, mental and physical health conditions, and positive impact on social relationships. (Section 4.1)
- Features serving the same types of rewards have varying effects on regret when the ways of presenting rewards are different, and vice versa. (Section 4.2)
- Habitual checking leaves a user with regret when it occurs repeatedly and fails to deliver new content, hence no reward. This mainly appears with the use of subscription-based features. (Section 4.3.1)
- Sidetracking from the original intention of use results in regret because of the delay in the expected, actual rewards of their intended use. (Section 4.3.2)
- Once landing on recommendation-based features, sometimes due to sidetracking, the infinite stream of recommended contents trigger users to stay on the app "just a bit more" and fall down the rabbit hole of regret. (Section 4.3.3)
- Feature-level summary of smartphone usage helped users (1) identify how much time was spent out of necessity, (2) prevent self-justification on usage by providing quantified evidences of unnecessary use, and (3) design an actionable plan on which part of app usage they should improve. (Section 4.4)

5 DISCUSSION

In this study, we introduce two novel approaches to understand smartphone use: (1) using regret as a theoretical construct and (2) collecting app feature-level usage data and experience samples to analyze intra-app usage behaviors and patterns. We discuss lessons learned from our feature-level analysis on regretful smartphone use, providing our insights on why people continue to repeat regretful use while being aware of it through the concept of delay discount and impulsive choice in literature. Lastly, we discuss design implications based on our findings and limitations of our study.

5.1 Why People Repeat Regretful Use: Delay Discount and Impulsive Choice

Regret theory in behavioral economics claim that when people make decisions, they consider the expected regret at the point of making the decision. In theory, our participants should have chosen to not use apps in a way they would regret. However, it is hard to break the repeated cycle of regretful use despite being aware of when and why they regret. We interpret why people continue repeating regretful use through the delay-discounting model of impulsiveness [5, 6, 26, 34, 42, 43, 45]. Delay-discounting is defined as "the reduction in the present value of a future reward as the delay to that reward increases"; and an impulsive choice refers to "the choice of a smaller, sooner reward over a larger, later rewards, when the larger reward would have been chosen at sufficiently longer delays of both rewards" [26].

In our context of smartphone use, the "smaller, sooner reward" would correspond to the four rewards (social activities, information, personal interests, and entertainment), while the "larger, later rewards" would be maintaining focus on ongoing tasks, advancing the moment of accomplishment, and moving up the future rewards such as better physical condition the next day and achieving goals of self-fulfillment. Examples of triggers to regretful experiences such as habitual use, deviation from original purpose, and getting dragged into a rabbit hole are results of making a series of impulsive choices.

It is important to note that some design components of the current social media apps are designed to encourage users to make impulsive choices, making it harder to self-regulate regretful uses, corresponding to the "dark patterns" [10, 49, 50]. One pattern we observed is placing attractive system recommended features in the middle of paths that a user must pass by (e.g., YouTube's

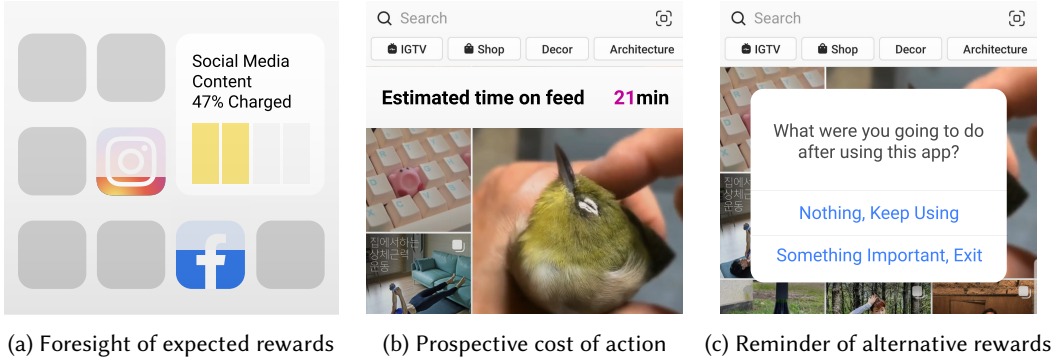(a) Foresight of expected rewards (b) Prospective cost of action (c) Reminder of alternative rewards

Fig. 6. Example wireframes portraying design implications for digital wellbeing based on our findings.

front page is full of recommended videos through reinforcement as shown in Appendix B.4's BROWSE_HOME). Another is making a seamless transition from necessary features to addictive features (e.g., Instagram recently connected SUGGESTED_POSTS at the end of the FOLLOWING_POSTS as shown in Appendix B.3's SUGGESTED_POSTS (a)). The variable schedule of rewards in content recommendations also play a huge role in hooking users.

Future designs for digital wellbeing should consider how to mitigate the inclination to impulsive choices and increase awareness of larger, later rewards through design for better self-regulation.

5.2 Design Implications

We leverage the concept of regret to analyze smartphone use behaviors at the app feature level. Regret is a retrospective measure that judges whether the actual rewards of their action exceeded those of other alternative choices. Acting as a proxy for what a user values, regret also entails the notion of whether the user wants to repeat the same action the next time they encounter a similar decision making situation. In our analysis, we exploited the retrospective trait of regret to identify which *parts* of an app session they would have rather not used. As a unit for the 'part,' we used *app features* as a proxy for different types and motivations of app use.

Our analyses on regretful smartphone use based on feature-level data suggest three design implications for digital wellbeing in social media app usage. The first two design implications aim to reduce regretful smartphone use, and the last one suggests a direction to strengthen self-reflection and self-monitoring.

5.2.1 Increasing users' foresight of the expected rewards of their action. The habitual checking pattern reveals that users are unconsciously seeking for rewards that are highly unlikely there when they open an app or switch to another feature during an app use. The main reason for regretting features that have a limited amount of contents at a time, such as Instagram's FOLLOWING_POSTS and VIEW_STORY and Facebook's NEWS_FEED, is the deficiency of contents and rewards due to repeated habitual checking. Participants made a false expectation that there will be newly updated contents (rewards) in the subscription-based features although there is a deficiency of rewards as a consequence of repeated checking. Therefore, it is important to empower users with better foresight of how much rewards are available to prevent repeated regret from content depletion.

An example design to strengthen a user's foresight would be to provide a preview on the status of newly available contents. Similar to a battery status visualization, apps could indicate the level of available contents through a widget with a phrase "Content x% charged" or visualization by coloring only x% of an app's icon to show how much content is 'filled up' (Figure 6a). We expect

that showing an approximation of how much rewards are available as a preview could minimize the gap between a user's (unconscious) expectation of rewards and actual available rewards. In this way, users might have a higher satisfaction for using an app that could increase the retention and sustainable use and thus benefit the app provider in the long run.

5.2.2 Enhancing users' awareness of the possible rewards of alternative actions. The patterns of deviating from the original intention and falling into a rabbit hole resulted in regret as these led to unexpectedly large time cost, thus proportionally increasing the possible alternative rewards had they spent the time on other things. Distraction from the originally intended use caused participants to unconsciously roam around other unintended features and regret later because their initial, desired rewards are delayed at higher cost. In addition, the false prediction of consuming "just a bit more" of time blinded the participants' awareness of how much time they have spent. The sidetracking and rabbit hole patterns make users become oblivious to the possible rewards of alternative activities to the current feature use, which are piling up as their feature usage gets longer. Thus, it is critical to remind users of how much time they have spent or will spend, and what rewards they could have earned otherwise.

The biggest reason of why people persist these patterns is distraction by immediate, bite-sized rewards presented up front in a feed. An example intervention to incapacitate the distraction could be reminding a user of the expected duration of their prospective usage when a user transitions to another feature. One's expected use time could be predicted based on the history of feature uses or a computational behavior model, similar to Medium's estimated read time [2], as shown in Figure 6b. Based on our findings, we pose that it might be more effective to remind the user of their *prospective* cost (i.e., how much more time they will spend) rather than their *retrospective* cost (i.e., how much time they have spent so far) as in traditional screen time reminders. This is because in the context of smartphone use, users lack conscious awareness of their cost and rewards and are easily deceived by immediate rewards. An intervention could further prompt a user with a question to remind about the alternative actions, "What were you doing before starting to use this app?" or "What were you going to do after using this?" (Figure 6c). Since such an intervention could be highly interrupting if it appears regardless of the usage context, it could be adjusted to appear when the rabbit hole pattern is detected, e.g., > 5 min stay or 10 videos/posts in Facebook's WATCH_VIDEO or Instagram's SUGGESTED_POSTS.

For users who want more coercive interventions, removing the source of distraction or limiting the number of content in recommendation-based features (e.g., three videos or ten posts) for a certain time period could also break the chain of "just a bit more" behaviors. In our interviews, some participants suggested that they even want to get rid of certain features, such as KakaoTalk's NEWS and Instagram's SUGGESTED_POSTS from the app if they could. P8 reported *"I wish I didn't have the news feature in KakaoTalk. It is a chat-oriented app, so I have to use it to communicate. But when I enter the chat, the tab next to it is the news, so it naturally makes me enter and I waste my time."* The desire to remove certain features in an app shares the goal with GreaseDroid's app patching [29] and existing web browser extensions such as Facebook news feed eradicator [40].

5.2.3 Providing a high-resolution portrayal of one's smartphone usage. Our new findings enabled by feature-level analysis underline the need for more fine-grained dimensions of app usage with a high-resolution portrayal of one's smartphone usage. Section 4 shows that the number of partially regretful sessions (648) were 1.5× the number of fully regretful sessions (431). This result strongly suggests that user experiences vary within an app. Our qualitative findings also highlight how different features have varying associations with perceived rewards and regret. Our findings support that behavior analysis for smartphone uses should take a more fine-grained approach than the current app-level approach.

We looked at the dimension of features, and there is a wide array of opportunities to inspect more fine-grained usage behaviors. For example, we let participants select a feature segment(s) that they regret using, and some said they wanted to select specific videos among a series of YouTube videos they watched in a session. Future work remains to collect even finer-grained usage logs and ESM responses by unit-based segmentation of social media contents (e.g., videos, posts, and stories) or time-based segmentation to capture diverging experiences within the same feature.

In addition, our results show that regretful experiences differed by the characteristics of content and features. Instead of aggregating content with different characteristics into the same WATCH_VIDEO feature in YouTube, tracking the source of the content such as from BROWSE_HOME, SEARCH, or SUBSCRIPTIONS would allow more profound understandings on smartphone use. *Finesse* currently runs on a rule-based mechanism to detect features. Future work could advance the mechanism to be more generalizable and robust to app updates by applying a deep learning-based model to detect UI elements of mobile apps [51].

Our implementation of feature-level detection and display back to users in an ESM system has potential applications beyond digital wellbeing. As described by the Human Screenome Project [44], social media negatively influences diverse aspects of individuals and the society, e.g., "mental and physical health, education, relationships, even on politics and democracy." However, analyzing the impact on human behavior is challenging for the lack of methods to reveal "exactly what people are seeing and doing on their screens" without invading user privacy. Finesse alleviates this concern by enabling feature-level logging of smartphone and social media usage. Feature-level usage detection has further applications, for example, to analyze which features of an education app, game, or a shopping app are most engaging, to facilitate the investigation on what people are *doing* on their screens. In addition, behavioral researchers can utilize our ESM method based on the feature-level detection in their study, e.g., to trigger ESM only upon the use of certain features. In line with the effort to enable high-resolution screen behavior data, we intend to publicize the anonymized dataset and system diagram of our ESM tool for others to build upon this work.

5.3 Limitations

Our method has a few limitations. First, the definition and degree of regret would vary across individuals, but we asked a binary question, "Do you regret or not?" Although it was necessary to ask this question in a binary format to minimize users' response burden, ratings of the regret or other scaled responses would be needed for further investigation of regret. Moreover, in an effort to minimize the burden, our ESM did not include a question for users' context (e.g. current tasks). Instead, in the retrospective interview, we provided participants the date, time, feature usage timeline, and regretted instances to facilitate the recall of the usage context. With these information provided, we asked why they regretted using the selected instances and what they could have done instead not to regret. We report our findings about the contexts of social media use and related rewards in Section 4.1.2.

We also collected samples of only immediate regret right after each app use session. However, regret could vary over time because the subject of comparison (i.e., what might have happened if an individual had not used the phone) would vary over time. For instance, an individual might not regret watching very interesting videos on YouTube right after use, but they might regret after a few hours for not finishing their assignment on time. Further investigation is needed on how regretful experience changes by the timing of sampling.

Furthermore, our user study was conducted on a limited pool of participants with similar demographics, students at one of top universities during the semester. Our results therefore might be biased towards students who are sensitive to academic achievements and thus feel more regretful for the feeling of wasting time and not being productive enough. In addition, smartphone users in

different age groups or cultural backgrounds might appreciate different types of gratifications from phone use and experience different patterns of regretful use.

To reduce user burden to answer the ESM response and minimize its effect to serve as an intervention, we devised our time condition and probability-based sampling policy. Therefore, our data on regretful use might have a sampling bias resulting from our sampling conditions. That said, collecting high-quality ESM data is about balancing participants' data capture burden and researchers' data collection needs. Given the high response rate of 95.8%, we believe that our efforts to reduce the data capture burden paid off and that our data correctly represent subjects' typical smartphone usage behaviors.

Lastly, this work looks into the usage of social media apps on smartphones only. A user's feature usage patterns might vary across devices such as tablets and laptops. Future work could scale investigation on feature-level usage patterns to other devices.

6 CONCLUSION

We developed a feature-level smartphone use tracking application, *Finesse*. We conducted a one-week ESM study followed by a retrospective semi-structured interview to understand how app features are associated with the feeling of regret about smartphone use. We showed how four types of rewards are considered in the feeling of regret. We further revealed three regretful usage patterns: (1) content depletion due to frequent checking habits, (2) feature tour and unconscious deviation from original motivation, and (3) falling down the rabbit hole in an infinite stream of bite-sized contents. Our findings unveiled through feature-level analysis emphasize the importance of studying an app's internal mechanisms and associated behaviors beyond traditional phone- and app-level analysis. Combined with feature-level analysis, the construct of regret acts as an effective language for communicating users' preferences on which parts of smartphone use they do not want to repeat and why. The nuanced understanding of regretful smartphone use offers new insights into social media app designs and intervention designs for digital well-being.

ACKNOWLEDGMENTS

REFERENCES

[1] [n.d.]. Digital Wellbeing. https://www.android.com/digital-wellbeing/
[2] [n.d.]. Read time. https://help.medium.com/hc/en-us/articles/214991667-Read-time
[3] 2020. Use Screen Time on your iPhone, iPad, or iPod touch. https://support.apple.com/en-us/HT208982
[4] 2021. Social Media Fact Sheet. https://www.pewresearch.org/internet/fact-sheet/social-media/?menuItem=3814afe3-3f3c-4623-910b-8a6a37885ab8
[5] George Ainslie. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychological bulletin* 82, 4 (1975), 463.
[6] George Ainslie and Richard J Herrnstein. 1981. Preference reversal and delayed reinforcement. *Animal Learning & Behavior* 9, 4 (1981), 476–482.
[7] Betsy Atkins. 2019. Are Super Apps The Future? *Forbes* (Sept. 2019). https://www.forbes.com/sites/betsyatkins/2019/09/03/are-super-apps-the-future/?sh=6cfd39896fd5
[8] Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: the short of mobile device use sessions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14.* ACM Press, Toronto, ON, Canada, 243–252. https://doi.org/10.1145/2628363.2628380
[9] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, Harris Cooper, Paul M. Camic, Debra L. Long, A. T. Panter, David Rindskopf, and Kenneth J. Sher (Eds.). American Psychological Association, Washington, 57–71. https://doi.org/10.1037/13620-004
[10] Harry Brignull. [n.d.]. Dark Patterns. https://darkpatterns.org

[11] Yunhwan Chae. 2020. S. Koreans spent nearly 30 hours on YouTube in Sept.: data. *Yonhap News Agency* (Oct. 2020). https://en.yna.co.kr/view/AEN20201008003900320

[12] J. Clement. 2019. *Most popular mobile social networking apps in the United States as of September 2019, by average session duration.* Technical Report. Statista, United States. https://www.statista.com/statistics/579411/top-us-social-networking-apps-ranked-by-session-length/

[13] Kakao Corporation. [n.d.]. KakaoTalk. https://play.google.com/store/apps/details?id=com.kakao.talk

[14] Shirley Cramer and Becky Inkster. 2017. . https://www.rsph.org.uk/our-work/campaigns/status-of-mind.html

[15] Android Developers. 2021. android.accessibilityservice. https://developer.android.com/reference/kotlin/android/accessibilityservice/package-summary

[16] Nir Eyal. 2020. Want to Design User Behavior? Pass the 'Regret Test' First. https://www.nirandfar.com/regret-test/

[17] Barney G Glaser and Anselm L Strauss. 1967. *Discovery of grounded theory: Strategies for qualitative research.* Routledge.

[18] Alexis Hiniker, Sungsoo (Ray) Hong, Tadayoshi Kohno, and Julie A. Kientz. 2016. MyTime: Designing and Evaluating an Intervention for Smartphone Non-Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, San Jose California USA, 4746–4757. https://doi.org/10.1145/2858036.2858403

[19] Alexis Hiniker, Shwetak N. Patel, Tadayoshi Kohno, and Julie A. Kientz. 2016. Why would you do that? predicting the uses and gratifications behind smartphone-usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, Heidelberg Germany, 634–645. https://doi.org/10.1145/2971648.2971762

[20] Daniel Hintze, Philipp Hintze, Rainhard D. Findling, and René Mayrhofer. 2017. A Large-Scale, Long-Term Analysis of Mobile Device Usage Characteristics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (June 2017), 1–21. https://doi.org/10.1145/3090078

[21] Daniel Kahneman and Amos Tversky. 1982. The simulation heuristic. In *Judgment under Uncertainty* (1 ed.), Daniel Kahneman, Paul Slovic, and Amos Tversky (Eds.). Cambridge University Press, 201–208. https://doi.org/10.1017/CBO9780511809477.015

[22] Colin Yoon-seung Kang. 2018. KakaoTalk continues to dominate mobile messenger market. http://yna.kr/AEN20180627005100320

[23] Jaejeung Kim, Chiwoo Cho, and Uichin Lee. 2017. Technology Supported Behavior Restriction for Mitigating Self-Interruptions in Multi-device Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 1–21. https://doi.org/10.1145/3130932

[24] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (March 2019), 1–29. https://doi.org/10.1145/3314403

[25] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.* ACM Press, Glasgow, Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300927

[26] Kris N Kirby, Nancy M Petry, and Warren K Bickel. [n.d.]. Heroin Addicts Have Higher Discount Rates for DelayedRewards Than Non-Drug-Using Controls. ([n. d.]), 10.

[27] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n' LoL: Group-based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, San Jose California USA, 998–1010. https://doi.org/10.1145/2858036.2858568

[28] Minsam Ko, Kyong-Mee Chung, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, and Junehwa Song. 2015. NUGU: A Group-based Intervention App for Improving Self-Regulation of Limiting Smartphone Use. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15.* ACM Press, Vancouver, BC, Canada, 1235–1245. https://doi.org/10.1145/2675133.2675244

[29] Konrad Kollnig, Siddhartha Datta, and Max Van Kleek. 2021. I Want My App That Way: Reclaiming Sovereignty Over Personal Devices. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–8.

[30] Geza Kovacs, Zhengxuan Wu, and Michael S. Bernstein. 2018. Rotating Online Behavior Change Interventions Increases Effectiveness But Also Increases Attrition. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–25. https://doi.org/10.1145/3274364

[31] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. https://doi.org/10.18637/jss.v082.i13

[32] Simone Lanette, Phoebe K. Chua, Gillian Hayes, and Melissa Mazmanian. 2018. How Much is 'Too Much'?: The Role of a Smartphone Addiction Narrative in Individuals' Experience of Use. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–22. https://doi.org/10.1145/3274370

[33] Janette Lehmann, Mounia Lalmas, Elad Yom-Tov, and Georges Dupret. 2012. Models of User Engagement. In *User Modeling, Adaptation, and Personalization*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu

Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Judith Masthoff, Bamshad Mobasher, Michel C. Desmarais, and Roger Nkambou (Eds.). Vol. 7379. Springer Berlin Heidelberg, Berlin, Heidelberg, 164–175. https://doi.org/10.1007/978-3-642-31454-4_14 Series Title: Lecture Notes in Computer Science.

[34] Alexandra W Logue. 1988. Research on self-control: An integrating framework. *Behavioral and Brain Sciences* 11, 4 (1988), 665–679.

[35] Graham Loomes and Robert Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal* 92, 368 (Dec. 1982), 805. https://doi.org/10.2307/2232669

[36] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the Design of YouTube Influences User Sense of Agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* https://doi.org/10.1145/3411764.3445467

[37] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26. https://doi.org/10.1145/3191754

[38] Ulrik Lyngs, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. 2018. "So, Tell Me What Users Want, What They Really, *Really* Want!". In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, Montreal QC Canada, 1–10. https://doi.org/10.1145/3170427.3188397

[39] Ulrik Lyngs, Kai Lukoff, Petr Slovak, Reuben Binns, Adam Slack, Michael Inzlicht, Max Van Kleek, and Nigel Shadbolt. 2019. Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.* ACM Press, Glasgow, Scotland Uk, 1–18. https://doi.org/10.1145/3290605.3300361

[40] Ulrik Lyngs, Kai Lukoff, Petr Slovak, William Seymour, Helena Webb, Marina Jirotka, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2020. 'I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20).* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376672

[41] Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (Jan. 2012), 105–114. https://doi.org/10.1007/s00779-011-0412-2

[42] Howard Rachlin. 1989. *Judgment, decision, and choice: A cognitive/behavioral synthesis.* WH Freeman/Times Books/Henry Holt & Co.

[43] Howard Rachlin and Leonard Green. 1972. Commitment, choice and self-control 1. *Journal of the experimental analysis of behavior* 17, 1 (1972), 15–22.

[44] Byron Reeves, Thomas Robinson, and Nilam Ram. 2020. Time for the Human Screenome Project. *Nature* 577 (Jan 2020), 314–317. https://doi.org/10.1038/d41586-020-00032-5

[45] Jerry B. Richards, Lan Zhang, Suzanne H. Mitchell, and Harriet de Wit. 1999. DELAY OR PROBABILITY DISCOUNTING IN A MODEL OF IMPULSIVE BEHAVIOR: EFFECT OF ALCOHOL. *Journal of the Experimental Analysis of Behavior* 71, 2 (March 1999), 121–143. https://doi.org/10.1901/jeab.1999.71-121

[46] John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2016. Personal Tracking of Screen Time on Digital Devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, San Jose California USA, 284–296. https://doi.org/10.1145/2858036.2858055

[47] Colleen Saffrey, Amy Summerville, and Neal J. Roese. 2008. Praise for regret: People value regret above other negative emotions. *Motivation and Emotion* 32, 1 (March 2008), 46–54. https://doi.org/10.1007/s11031-008-9082-4

[48] Choonsung Shin and Anind K. Dey. 2013. Automatically detecting problematic use of smartphones. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13.* ACM Press, Zurich, Switzerland, 335. https://doi.org/10.1145/2493432.2493443

[49] Jonathan A. Tran, Katie S. Yang, Katie Davis, and Alexis Hiniker. 2019. Modeling the Engagement-Disengagement Cycle of Compulsive Phone Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19.* ACM Press, Glasgow, Scotland Uk, 1–14. https://doi.org/10.1145/3290605.3300542

[50] José P Zagal, Staffan Björk, and Chris Lewis. 2013. Dark patterns in the design of games. In *Foundations of Digital Games 2013.*

[51] Xiaoyi Zhang, Lilian de Greef, Amanda Swearngin, Samuel White, Kyle Murray, Lisa Yu, Qi Shan, Jeffrey Nichols, Jason Wu, Chris Fleizach, Aaron Everitt, and Jeffrey P. Bigham. 2021. Screen Recognition: Creating Accessibility Metadata for Mobile Applications from Pixels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).* https://doi.org/10.1145/3411764.3445186

A FEATURE LIST

A.1 KakaoTalk

Feature name	Description	User Action	Form of Content	Source of Content
CHAT	chatting with others	chatting		
IN_CHAT_SEARCH	doing web search inside a chat room	searching		
CONTACTS	viewing others' profile photo and information	viewing	profile	
NEWS	reading news recommended by KakaoTalk	viewing	news feed	recommendation-based
ETC	viewing a tab with various sub-features such as payment, emoticon shops, etc.	viewing	menu	

A.2 Facebook

Feature name	Description	User Action	Form of Content	Source of Content
MESSENGER	chatting with other users using Facebook messenger	chatting		
SEARCH	searching accounts, tags, posts, etc.	searching		
UPLOAD_POST	uploading Facebook post	uploading	post	
NOTIFICATIONS	checking notification	viewing	notification	
MENU	exploring Facebook menu	viewing	menu	
COMMENTS	viewing comments of a post	viewing	comment	
IN_APP_WEB	using Facebook web browser	viewing	web	
VIEW_POST	viewing a post	viewing	post	
NEWS_FEED	viewing the Facebook news feed	viewing	post feed	subscription-based
GROUPS	viewing the posts from a group	viewing	post feed	subscription-based (groups only)
MY_TIMELINE	viewing profile or posts of the user	viewing	profile	created by the user
OTHER'S_TIMELINE	viewing profile or posts of other users	viewing	profile	other users
PAGE'S_TIMELINE	viewing the posts from a page	viewing	profile	page
WATCH_VIDEO	watching videos	viewing	video	recommendation-based
VIEW_STORY	viewing Facebook story	viewing	story	subscription-based

A.3 Instagram

Feature name	Description	User Action	Form of Content	Source of Content
SEARCH	searching accounts, tags, places, etc.	searching		
DIRECT_MESSAGE	chatting with other users using direct message	chatting		
UPLOAD_POST	uploading Instagram post	uploading	post	
UPLOAD_STORY	uploading Instagram story	uploading	story	
VIEW_STORY	viewing Instagram story	viewing	story	subscription-based
NOTIFICATIONS	checking notification	viewing	notification	
IN_APP_WEB	using Instagram web browser	viewing	web	
WATCH_VIDEO	viewing videos	viewing	video feed	recommendation-based
FOLLOWING_POSTS	viewing the feed with posts of the following users	viewing	post feed	subscription-based
SUGGESTED_POSTS	viewing the feed with posts recommended by Instagram	viewing	post feed	recommendation-based
VIEW_BY_HASHTAG	viewing posts by hashtags	viewing	post feed	filtered by hashtag
MY_POSTS	viewing the profile and posts of the user	viewing	profile	created by the user
OTHER'S_POSTS	viewing profile or posts of other users	viewing	profile	other users

A.4 Youtube

Feature name	Description	User Action	Form of Content	Source of Content
SEARCH	searching videos	searching		
NOTIFICATIONS	checking notification	viewing	notification	
WATCH_VIDEO	watching videos	viewing	video	
COMMENTS	reading comments	viewing	comment	
CHANNEL_PAGE	viewing a channel's page	viewing	profile	
BROWSE_HOME	viewing home tab with recommended videos	viewing	feed	recommendation-based
EXPLORE	viewing explore tab with trending videos	viewing	feed	recommendation-based, trending
SUBSCRIPTIONS	viewing subscriptions tab with subscribing channels' videos	viewing	feed	subscription-based
PLAYLISTS	exploring playlists	viewing	playlist	recommendation-based
MY_LIBRARY	viewing library tab with recent, history, etc.	viewing	playlist	created by the user

B FEATURE SCREENSHOTS

B.1 KakaoTalk
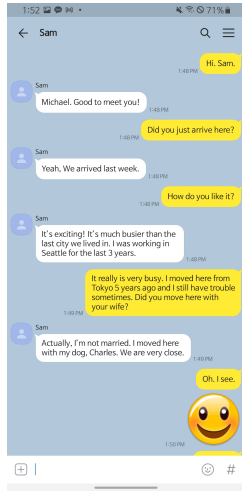
CHAT   IN_CHAT_SEARCH (a)   IN_CHAT_SEARCH (b)
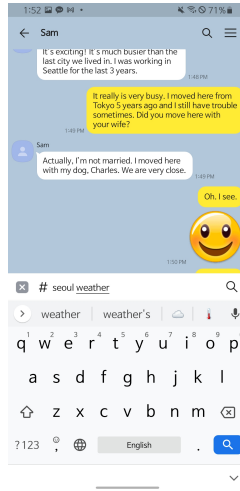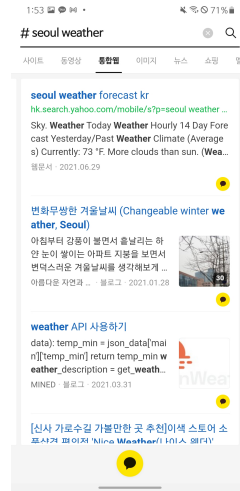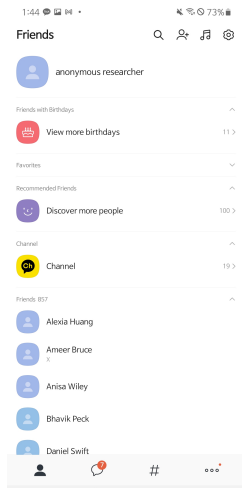
CONTACTS (a)   CONTACTS (b)   NEWS   ETC
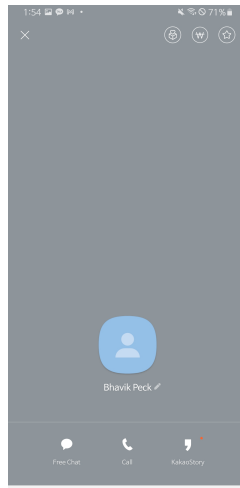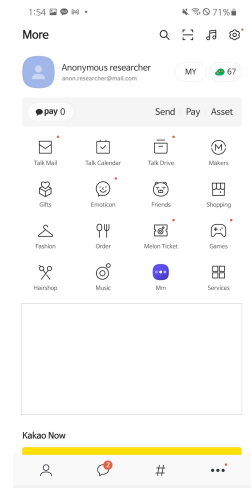
B.2 Facebook

MESSENGER
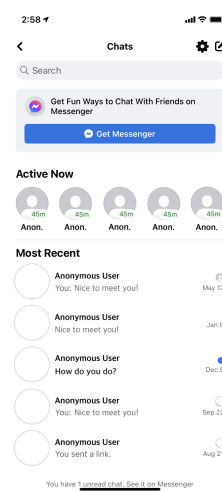
SEARCH
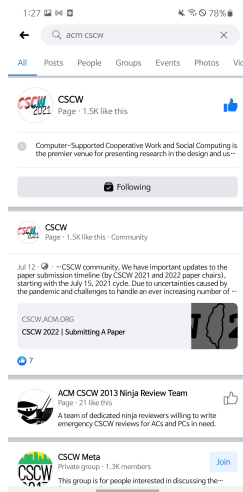
UPLOAD_POST
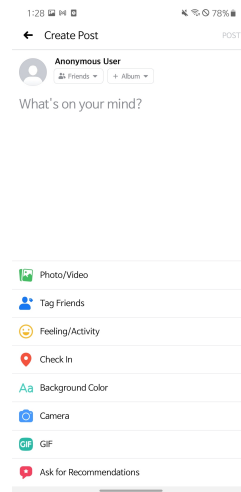
NOTIFICATIONS
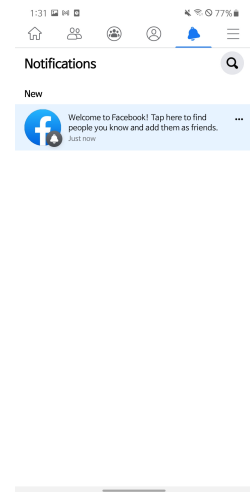
MENU
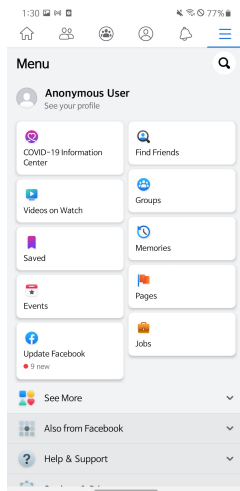
COMMENTS
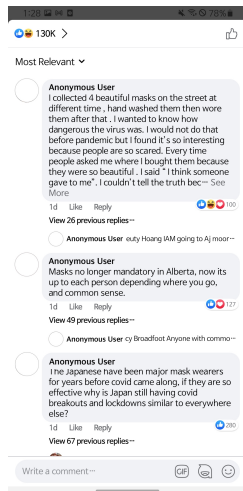
IN_APP_WEB

VIEW_POST
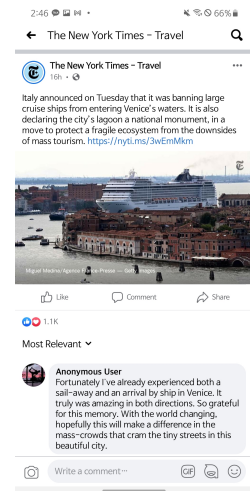
NEWS_FEED
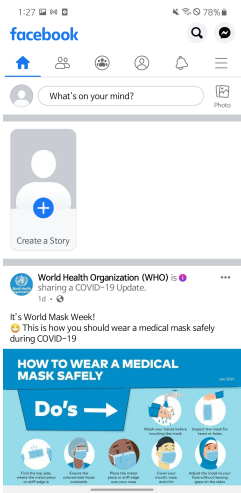
GROUPS
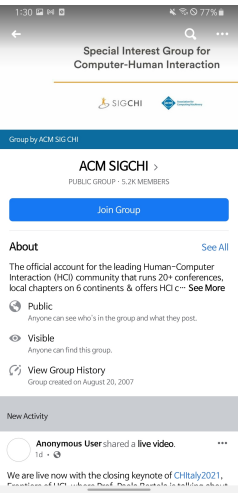
MY_TIMELINE
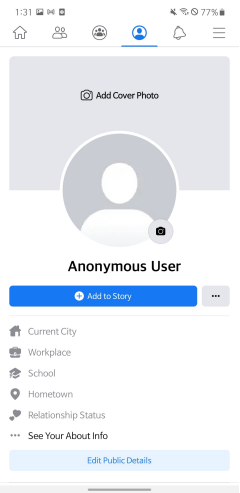
OTHER'S_TIMELINE

PAGE'S_TIMELINE
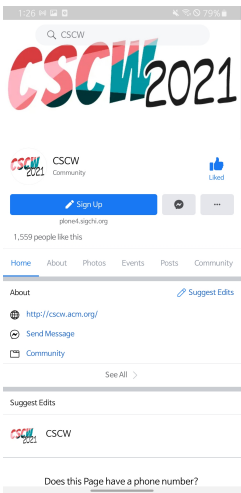
WATCH_VIDEO

VIEW_STORY
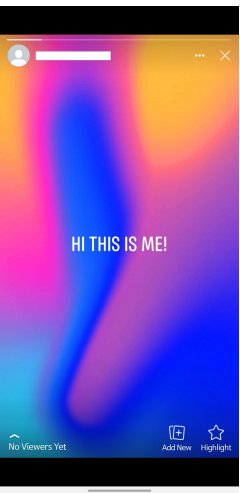
B.3 Instagram

SEARCH
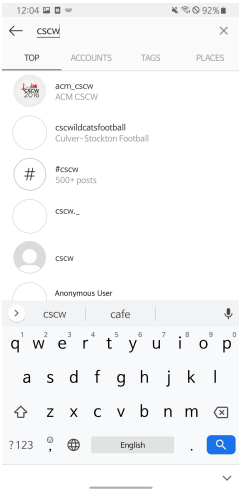
DIRECT_MESSAGE
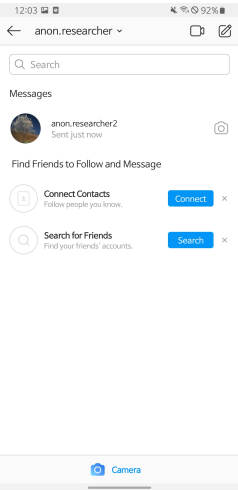
UPLOAD_POST (a)
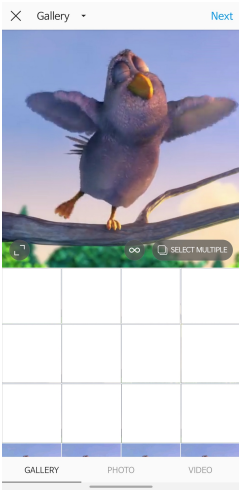
UPLOAD_POST (b)
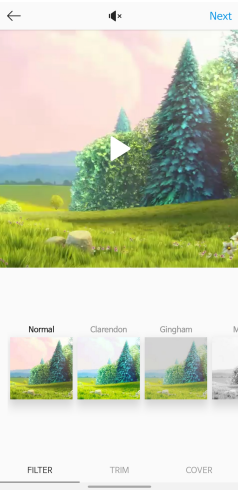
UPLOAD_POST (c)
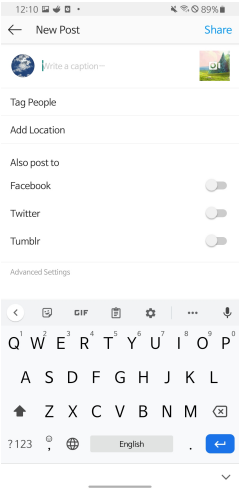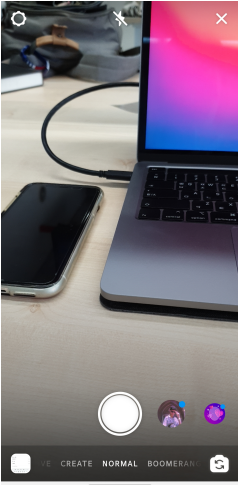
UPLOAD_STORY (a)   UPLOAD_STORY (b)   UPLOAD_STORY (c)   UPLOAD_STORY (d)
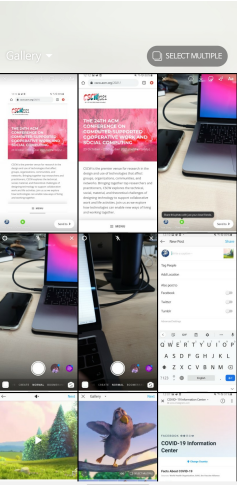
VIEW_STORY   NOTIFICATIONS   IN_APP_WEB   WATCH_VIDEO
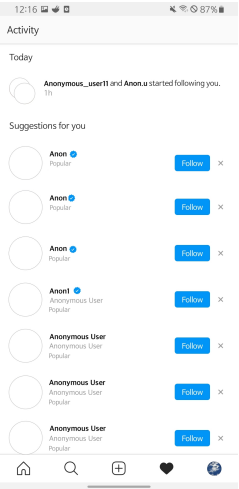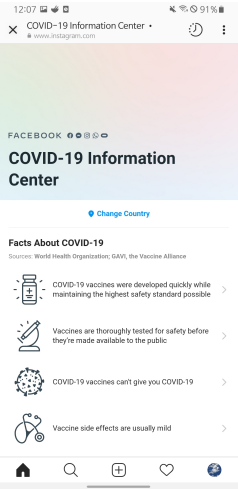
FOLLOWING_POSTS
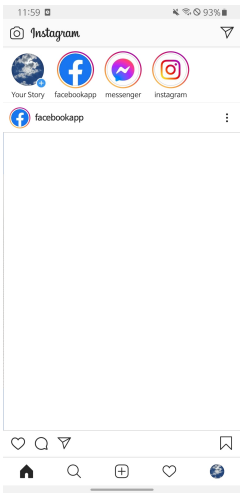
SUGGESTED_POSTS (a)
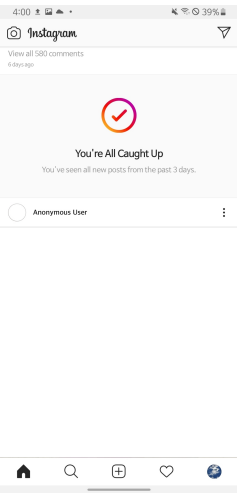
SUGGESTED_POSTS (b)
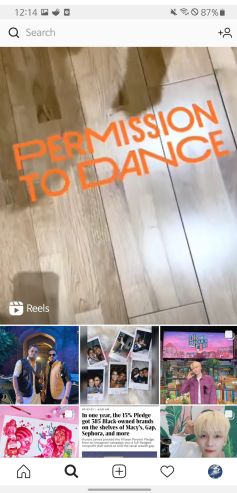
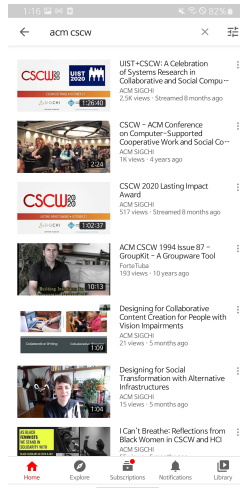SUGGESTED_POSTS (c)

VIEW_BY_HASHTAG
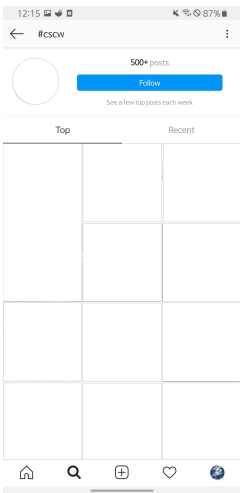
MY_POSTS
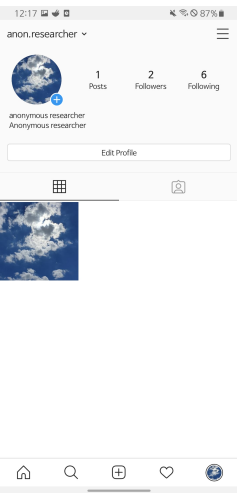
OTHER'S_POSTS
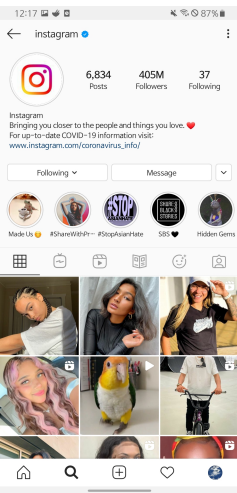
B.4 YouTube

SEARCH (a)
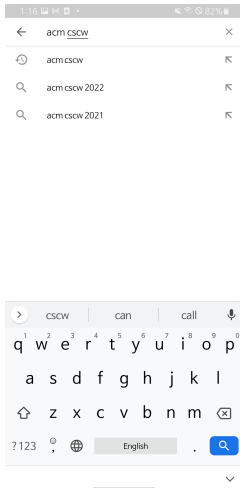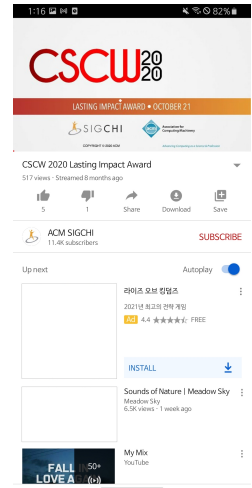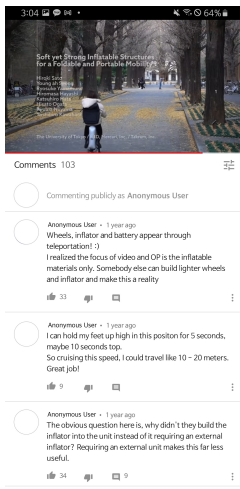

SEARCH (b)

NOTIFICATIONS

WATCH_VIDEO
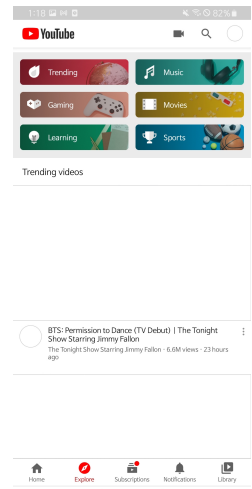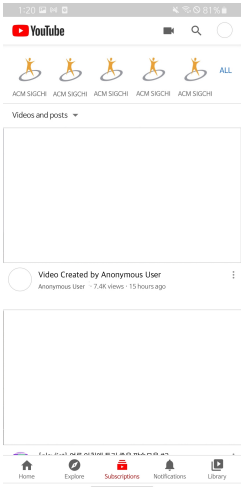

COMMENTS

CHANNEL_PAGE
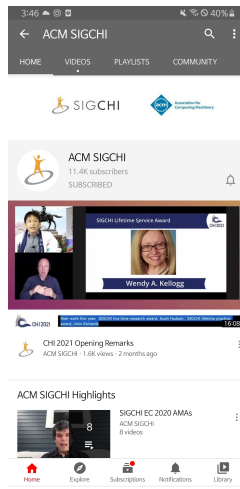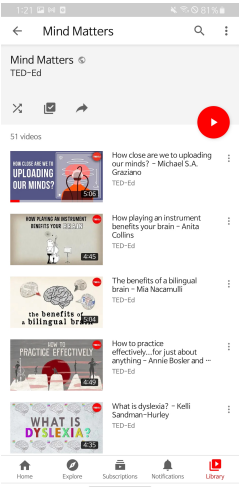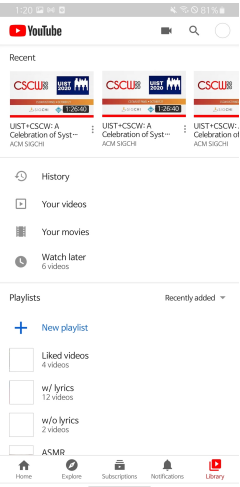

BROWSE_HOME
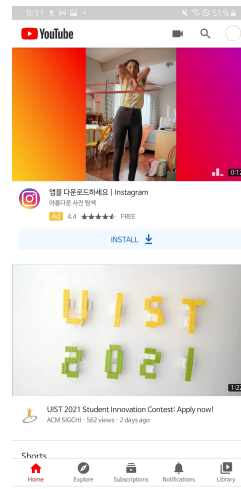
EXPLORE

SUBSCRIPTIONS

PLAYLISTS

MY_LIBRARY

C SUMMARY OF GENERAL LINEAR MIXED MODEL ANALYSIS

Feature	Estimate	Std. Error	z value	p-value		Median Regret Ratio
Instagram (Baseline: FOLLOWING_POSTS)						0.17
(Intercept)	-1.79	0.24	-7.33	0.00	***	
SUGGESTED_POSTS	0.94	0.15	6.11	0.00	***	0.31
VIEW_BY_HASHTAG	0.63	0.53	1.19	0.23		0.0
VIEW_STORY	0.30	0.11	2.80	0.01	**	0.18
OTHER'S_POSTS	0.03	0.14	0.23	0.82		0.13
IN_APP_WEB	-0.08	0.70	-0.11	0.91		0.0
UPLOAD_STORY	-0.22	0.27	-0.82	0.41		0.0
WATCH_VIDEO	-0.34	0.37	-0.92	0.36		0.0
SEARCH	-0.38	0.29	-1.31	0.19		0.0
NOTIFICATIONS	-1.13	0.22	-5.09	0.00	***	0.0
DIRECT_MESSAGE	-1.20	0.20	-5.92	0.00	***	0.0
UPLOAD_POST	-1.60	0.64	-2.50	0.01	*	0.0
MY_POSTS	-1.71	0.21	-8.08	0.00	***	0.0
Facebook (Baseline: NEWS_FEED)						0.26
(Intercept)	-1.29	0.33	-3.97	0.00	***	
WATCH_VIDEO	0.32	0.27	1.18	0.24		0.14
IN_APP_WEB	-0.33	0.49	-0.67	0.50		0.0
SEARCH	-0.45	0.43	-1.06	0.29		0.0
OTHER'S_TIMELINE	-0.55	0.45	-1.23	0.22		0.0
GROUPS	-0.66	1.26	-0.52	0.60		0.0
MY_TIMELINE	-0.98	0.92	-1.06	0.29		0.33
PAGE'S_TIMELINE	-1.20	0.45	-2.66	0.01	**	0.0
NOTIFICATIONS	-1.41	0.41	-3.40	0.00	***	0.0
MENU	-2.34	1.11	-2.11	0.03	*	0.0
VIEW_STORY	-13.93	611.82	-0.02	0.98		0.0
UPLOAD_POST	-14.57	878.92	-0.02	0.99		0.0
YouTube (Baseline: BROWSE_HOME)						0.04
(Intercept)	-2.60	0.32	-8.14	0.00	***	
NOTIFICATIONS	0.97	0.98	0.99	0.32		0.0
EXPLORE	0.87	0.72	1.21	0.23		0.0
COMMENTS	0.35	0.49	0.70	0.48		0.42
SUBSCRIPTIONS	-0.05	0.34	-0.14	0.88		0.0
WATCH_VIDEO	-0.08	0.14	-0.60	0.55		0.07
MY_LIBRARY	-0.61	0.76	-0.81	0.42		0.0
SEARCH	-0.93	0.33	-2.84	0.00	**	0.0
PLAYLISTS	-14.93	2622.91	-0.01	1.00		0.0
CHANNEL_PAGE	-16.08	2347.32	-0.01	0.99		0.0
KakaoTalk (Baseline: CHAT)						0.05
(Intercept)	-2.83	0.30	-9.49	0.00	***	
NEWS	2.83	0.34	8.22	0.00	***	0.12
IN_CHAT_SEARCH	1.30	0.54	2.41	0.02	*	0.0
CONTACTS	0.70	0.15	4.57	0.00	***	0.07
ETC	0.24	0.30	0.80	0.43		0.0

*$p<0.05$, **$p<0.01$, ***$p<0.001$

Table 2. Results of general linear mixed model (GLMM) analysis on features and regret. The table shows the estimated amount of increase or decrease in odds of regret (Estimate), its standard error (Std. Error), z-value of z-statistics (z-value), and the estimate's significance (p-value). For example, in the case of KakaoTalk, relative to CHAT, participants were 2.83x more likely to regret using NEWS; in the case of Instagram, 0.94x more likely to regret using SUGGESTED_POSTS than using FOLLOWING_POSTS. The rightmost column shows the median regret ratio among participants who used the feature for reference.