Sad or just jealous? Using Experience Sampling to Understand and Detect Negative Affective Experiences on Instagram

Mintra Ruensuk
UNIST
Ulsan, Republic of Korea
mintra@unist.ac.kr

Taewan Kim
KAIST
Daejeon, Republic of Korea
taewan@kaist.ac.kr

Hwajung Hong
KAIST
Daejeon, Republic of Korea
hwajung@kaist.ac.kr

Ian Oakley
UNIST
Ulsan, Republic of Korea
ian.r.oakley@gmail.com

ABSTRACT

Social Network Services (SNSs) evoke diverse affective experiences. While most are positive, many authors have documented both the negative emotions that can result from browsing SNS and their impact: Facebook depression is a common term for the more severe results. However, while the importance of the emotions experienced on SNSs is clear, methods to catalog them, and systems to detect them, are less well developed. Accordingly, this paper reports on two studies using a novel contextually triggered Experience Sampling Method to log surveys immediately after using Instagram, a popular image-based SNS, thus minimizing recall biases. The first study improves our understanding of the emotions experienced while using SNSs. It suggests that common negative experiences relate to appearance comparison and envy. The second study captures smartphone sensor data during Instagram sessions to detect these two emotions, ultimately achieving peak accuracies of 95.78% (binary appearance comparison) and 93.95% (binary envy).

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **Ubiquitous and mobile computing systems and tools**; *Smartphones.*

KEYWORDS

Affect detection, Social media, Smartphones, Experience sampling, Instagram

ACM Reference Format:
Mintra Ruensuk, Taewan Kim, Hwajung Hong, and Ian Oakley. 2022. Sad or just jealous? Using Experience Sampling to Understand and Detect Negative Affective Experiences on Instagram. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3517561

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517561

1 INTRODUCTION

Online Social Networking Services (SNS) allow users to construct a profile, create connections to other users' profiles and view and traverse their list of connections [10]. They have attracted billions of users. In 2020, the core products of Facebook (e.g., Instagram, Messenger), one of the major players in the area, attracted 2.80 billion monthly active users with 1.84 billion of these users visiting the services on a daily basis [32]. Indeed, recent estimates indicate 72% of American adults use at least one SNS, although different demographic groups favor different services. Facebook, for example, was used by 77% of adults (30-49 years old), while Instagram was most popular amongst young adults, with 71% of 18–29 year olds indicating they use the service [14]. SNSs attract users by enabling them to view, browse, post, and react to diverse media content including images [120], short videos [60], and live videos [91]. Posted content covers a vast range of topics and formats, such as comedy and pranks [58] or blogging and lifestyle (e.g., traveling [80] or exercise [56]) and a widely acknowledged part of the appeal of SNSs is the ability of the content to influence people's emotional states [42, 63, 74]. These affective effects are highly diverse and frequently positive. For example, at the more traumatic end of the scale, Facebook users may cope better with stress after being made unemployed [13]. More mundanely, spending time on Instagram may boost benign envy, a generally beneficial affective experience that can motivate self-development [78].

However, there is a darker side. A wide body of literature catalogues the ways in which SNS activity can negatively impact people's emotional states. These less desirable outcomes can be triggered by specific incidents, such as being un-friended on Facebook, an event that may result in rumination—a prolonged, repetitive process of dwelling or consciously thinking about a situation that is associated with anxiety and depression [8]. Problems can also occur during in more day-to-day use. For example, social comparison can lead to emotions such as jealousy, anxiety, and irritation [37]. These negative affective experiences are particularly problematic because they typically go unremarked—many users are unaware of situations in which SNS negatively affect their emotions and do not manage these experiences appropriately [37]. The results of this mismanagement may include general decreases in affective well-being or self-esteem [57] due to experiences of envy [114] or unflattering social comparison [100]. It can also lead to antisocial behaviors such as trolling [18] or flaming [1]. Indeed the pervasiveness of these negative experiences has minted words in the public lexicon [99]: "Facebook Depression", for example, has long been used to refer to a depression that develops when spending prolonged periods on social media sites [84]

The importance and impact of the affective experiences evoked by SNS use has spurred a wide range of work. A key strand seeks to capture, document and understand the emotions experienced during SNS use. Much of this work examines immediate responses. Lin and Utz [68], for example, document the emotional impact of reading friends' posts on Facebook, while de Vries et al. [25] examine the affective experiences elicited by viewing positive images from a stranger's feed on Instagram. Other authors have explored the affective processes that underlie SNS use. For example, Liu et al. [70] explain the mechanisms behind social comparison and both Verduyn et al. [114] and Faelens et al. [33] look at the overall impact of SNSs on affective well-being. This research has been enabled by a wide range of methods including online questionnaires [68], empirical studies [25, 48], and field studies [92, 114]. This latter approach, typically instantiated as the Experience Sampling Method (ESM) [64], is particularly valuable as it uses repeated administration of survey questions to capture snapshots of participant's day-to-day experiences over protracted periods of several weeks. This method is useful for revealing the subjective experiences of people in their natural environments [101] and can reduce reliance on the participants' long-term memory to reconstruct past events [111]. While the benefits of the approach are clear, ESM is not without its limitations. Specifically, we note that the frequency (e.g., three [118] to six [33] times per day) and fixed timing (e.g., in the evening) of survey administration can still lead to widely acknowledged recall biases [82, 112] that can impact the quality of the data captured. We argue this issue is particularly problematic for studies of the emotional impact of SNS use: affective experiences fade. Recall of these experiences may be particularly prone to self-report errors such as forgetting, mis-estimating the severity of an affective experience, or even misidentifying one: differentiating negative emotions is challenging in general and particularly so for key user SNS groups such as young adults [83, 85].

Furthermore, while ESM has been found to be a practical and efficient technique to support research, the user burden of self-report activities, such as recording videos of activities [102] or reporting engagement levels during task performance [36], limit its applicability to real world applications that do more than catalogue or observe affective experiences. We argue that while ESM can support effective studies, it is challenging to use the technique to design effective applications that can aid users in their day to day lives. One potential solution to this problem lies in an emerging body of research that utilizes passive sensor data collected from mobile smart devices to recognize and differentiate an individual's emotional states [79, 115, 118]. Researchers have shown that the advanced motion, touch, or camera sensors built into mobile smart devices can be used to classify the emotional states that occur during their use in a wide range of situations such as game play [38, 51], data entry [39, 115], general smartphone operation [118] and, SNS use [96]. While the diversity of this work suggests the effects are robust—that smart device sensors can pick up cues that are reliably associated with specific affective experiences—a major current limitation is that existing studies have been predominantly conducted in highly controlled or lab-based settings. Studies rely on artificially elicited emotions [79], take place in fixed locations [96, 115], involve performance of highly specific, predetermined tasks (such as Fitts' law tapping [79]) or are conducted over very limited periods of time [38, 96]. We argue that while current work shows the potential for mobile smart devices to detect emotions during SNS use, it remains unclear whether or not this will be effective in real world SNS use scenarios involving users operating their own devices at times and locations, and for durations, of their choosing in order to view and interact with contents on their own genuine SNS accounts.

This paper aims to address these two empirical issues through two closely related studies. In the first study, we seek to improve the granularity of the affective data captured during ESM studies of social media use. We achieve this in two ways: by developing a mobile phone application that captures ESM survey data directly after a user finishes an SNS use session and; by composing our ESM survey to capture a wide range of finely grained affective constructs that have been found to be associated with SNS use. These include: valence and arousal [11], social comparison [41, 106], appearance comparison [109], envy [108], self-esteem [31, 94], and depression [6]. By surveying a relevant set of affective constructs in a timely manner, this study advances knowledge about the range and scope of affective experiences that occur during the course of regular, day-to-day SNS browsing sessions. The results of this study indicate that the negative affective experiences that most frequently occur in response to SNS use relate to the constructs of appearance comparison and envy. Building on this finding, the second study deploys a revised sensor enabled version of our ESM survey application in order to explore whether or not the day to day affective experiences of SNS users can be detected automatically using data from a smart device's sensors. Specifically, this system and study combines capture of sensor data, in the form of motion [79, 118] and touch [40, 115], from within SNS browsing sessions with ESM self-reports of appearance comparison and envy recorded immediately after the same SNS browsing sessions. We use the self-reported data as binary labels and develop classifiers using the sensor data capable of predicting binary appearance comparison with an accuracy of up to 95.78% and binary envy with an accuracy of up to 93.95%.

This paper makes two key contributions. Firstly, we extend existing descriptions of the affective experiences elicited by SNS use with an ESM method and data that captures self-reports immediately after SNS use sessions, minimizing potential recall biases. This work highlights the relative prevalence of negative appearance comparison and envy. Second, we build and demonstrate a system running on commercially available mobile devices that is able to predict these negative affective experiences with binary accuracies of greater than 93.95%, a level of performance we believe is sufficient to enable real-world applications that can help users reflect on their emotions during SNS use and empower them to take actions (e.g., pausing a session predicted to be stressful) that can support more effective management of their affective experiences.

2 RELATED WORK

2.1 SNS Use and Affective Experiences

A growing body of research reports on the links between SNS use and affective experiences in everyday life [46, 78, 106, 108]. Researchers have explored the presence of affective experiences in SNS using both broad constructs such as valence and arousal [24, 96] and also more specifically targeted constructs such as social comparison [25]. For example, Hasan et al. [47] analyzed text messages to characterize affective experiences in Twitter along two dimensions (valence and arousal) and Preoţiuc-Pietro et al. [90] created a system that predicts valence and arousal using 2895 anonymized Facebook posts labeled by psychologically trained annotators. In terms of more specific affective constructs, de Vries et al. [25] investigated the emotional consequences of viewing strangers' positive posts on Instagram from a social comparison perspective, Fardouly and Vartanian [34] examined the relationship between Facebook usage and appearance comparison, and Verduyn et al. [114] demonstrated how passive Facebook usage leads to declines in affective well-being by increasing envy. In addition, other types of affective experiences have been investigated including self-esteem and depression. Examples include Martinez-Pecino and Garcia-Gavilán [75]'s work to analyze the influence of likes and self-esteem on Instagram, and Guntuku et al. [46]'s work to examine which attributes of profiles and posted images are associated with depression in Twitter users. While this body of work is comprehensive, we note that the vast majority of work focuses on demonstrating the relationship between SNS use and affective experiences [34, 54]. While this is valuable, we argue it tells an incomplete story. In this paper, we seek to complement the literature on how SNS use impacts affective experiences by examining how frequently different experiences occur in day to day use.

2.2 Quantifying Affective Experiences in SNS

Prior studies have used various techniques to quantify affective experiences in social media including questionnaires, empirical studies, and field studies. For example, in terms of questionnaires, Lin and Utz [68] study the emotional outcomes of reading a post on Facebook and examine the role of tie strength in predicting happiness and envy while Chow and Wan [21] investigate whether pre-existing proclivities such as a tendency to engage in Facebook social comparison and envy are associated with depressive symptoms. Perhaps the major disadvantage of using questionnaires to quantify SNS experiences is emotional memory bias [82]—a tendency to recall events that match our moods. Indeed, there is clear evidence that our recall of experiences on social media is far from perfect. Nontasil and Payne [82], for example, showed participants could recall only 30% of encountered threads after consuming content on Facebook for 10-15 minutes. To deal with these recall problems, prior researchers have explored alternative study methods. For example, de Vries et al. [25] conducted a lab study to examine the emotional consequences of viewing strangers' positive Instagram posts. In this controlled setting, they observed that viewing the posts decreased positive affect among individuals with high levels of social comparison orientation. Similarly, Helmut Appel and Gerlach [48] conducted a quasi-experimental online study and noted that depressed participants showed elevated levels of envy, especially after seeing an attractive Facebook profile. While these findings strengthen the evidence for ties between SNS use and particular affective states by sidestepping recall biases, formal studies can lack ecological validity. It can be hard to generalize from effects observed in the lab to emotions experienced in the wild.

The Experience Sampling Method (ESM) [64] is one potential solution to this problem. It has been widely employed to quantify affective experiences in SNS. The method involves signaling participants several times per day for several weeks to complete brief surveys [86]. The key strength of ESM is that it reduces participants' dependence on long-term memory to reconstruct past events or experiences. Also, data collection can be primed to particular events that are of interest [111]. Prior research has employed ESM to understand various emotional effects of using SNS such as affective well-being [23, 33, 114] and loneliness [92]. However, such work tends to alert participants to complete a survey at a fixed time of a day (e.g., five times per day between 10 AM and 12 AM [114] and six times per day [33, 45]). The validity of assessing the affective experiences in SNS at random or fixed time intervals is questionable; prior work reports recall is poor after periods as short as 15 minutes [82]. In this work, we seek to improve data quality in ESM surveys by designing a system to capture self-reports directly after a user finishes an SNS use session.

2.3 Detecting Affective Experiences using Smartphones

Researchers have used smartphones to detect the affective experiences of users in a wide range of domains (e.g., game play [38, 51], mental health [53], social media [96]) by examining the various sensor data these devices can capture. Examples include data about speech patterns [16], physical activity [98], social communication [66], and phone calls [67]. However, some forms of smartphone data may expose a user's identification or other private or personal material. For example, Niforatos and Karapanos [81]'s work collected and analyzed 2953 self-face images to infer users' emotions while interacting with different categories of mobile apps (such as productivity and entertainment). Other researchers have used less explicit sensor channels that create fewer privacy issues while still showing strong capabilities to detect affective experiences. For example, Gao et al. [38]'s work used mobile touch data to detect user's emotional states while playing Fruit Ninja, a game based around rapid repeated stroke gestures, while Mottelson and Hornbæk [79]'s work used both mobile touch and motion data to estimate affect in the wild. In work highly relevant to this paper, Ruensuk et al. [96] demonstrated the feasibility of using mobile sensor data (including touch and motion) to classify a user's affective experience while using Facebook in a controlled setting with promising recognition performance of up to 94.16% for binary valence. In this work, we seek to apply the beneficial properties of sensor-based affect detection in a previously unexplored scenario—during genuine real world SNS use.

3 STUDY 1: UNDERSTANDING AFFECTIVE EXPERIENCES ON SNS

This study sought to capture the affective experiences that occur during real world SNS use at a high level of granularity in terms of both the range of affective constructs surveyed and the specificity with which we target genuine SNS use. To achieve this goal, we conducted a 14-day ESM study to capture the experiences of participants directly after consuming or responding to content on Instagram on their smartphones. Participants were requested to provide self-assessment of both usage behaviors and their affective states (e.g., appearance comparison [109], social comparison [41], self-esteem [94]) after using Instagram multiple times a day. The study closed with semi-structured interviews that sought to contextualize the quantitative results. The study was approved by the host university's Institutional Review Board (IRB).

3.1 Study Design

The study sought to quantify emotional states in real-world SNS use by deploying Experience Sampling Methods (ESM) [64]. Based on recent recommendations [49] and surveys [111] of the ESM literature, we selected a two week study period. A key aspect of ESM study design is the notification mechanism [65]—the scheme by which surveys are presented to participants [111]. We selected an event-contingent design [5] that prompts users immediately after closure of Instagram, providing their browsing session was sufficiently long (defined as longer than 15 seconds [72]) to likely involve content consumption or explicit interaction. To ensure surveys were not completed *post-hoc*, after some arbitrary delay, notifications presented to participants expired after five minutes and were no longer accessible. This was intended to reduce recall biases. Finally, to prevent participant fatigue, we configured notifications to occur at least 90 minutes apart. To implement this study protocol, we developed *IG Use*, an Android app that runs in the background and continuously monitors foreground use of the Instagram app [110]. To deal with any failures with this service, we implemented a system level alarm to run IG Use on phone startup and check for its presence at two-hour intervals. In this way we were able to reliably log data throughout the course of the two week study. IG Use logged Instagram use data (e.g., timestamp, status), notification log data (e.g., alert timestamp, status) and also presented and logged study survey data. All data was pushed to an online database, either immediately, or the next time the participant's phone connected to the Internet.

3.2 Materials

During the ESM study, participants repeatedly completed self assessments of their affective states. These took the form of short surveys targeting seven different constructs. Figure 1 illustrates the interface for completing these surveys in the IG Use app. The surveys were always presented in the following order and entailed:

- **Valence and Arousal.** We deployed the Self-Assessment Manikin (SAM) [11] to capture various degrees of emotional valence and arousal. SAM shows sets of pictorial manikins that depict emotions via facial expressions and bodily reactions (see Figure 1b). The valence dimension is represented as a frowning to a smiling face, while arousal features manikins in states between calm and highly tense or excited [11]. Participants select the manikins that most represent their emotional states on a 9-point numeric scale. We include valance and arousal due to their prevalence in the literature on smartphone affect detection [79, 103, 115].
- **Type of communication.** We adapted an existing scale to quantify the type of communication (either active or passive) prevalent during Instagram sessions [72]. We modified the question to ask specifically about Instagram (see Figure 1c). This construct is relevant as prior work has suggested that negative affect is more closely related to passive use of SNS (e.g., viewing, browsing) than active use (e.g., posting, commenting) [114].
- **Social comparison**. We followed prior work studying inspiration on Instagram [78] and measured the intensity of social comparison. We achieved this by adapting two items from the Facebook Social Comparison Scale [41, 106]. We modified the questions to enquire about emotional states after using Instagram and to specify the target of comparison. Figure 1d displays the two questions in a 5-point Likert scale.
- **Appearance comparison**. We used the state appearance comparison scale [109] to assess the amount Instagram users engage in appearance processing and draw comparisons with other Instagram users [35]. This involves two questions, each with a 5-point Likert scale.
- **Envy**. We adapted the Facebook envy scale [108] to capture envy experienced during SNS use [69]. We dropped two scale items that were item reversals and used a 5-point Likert scale for the remaining six items.
- **Self-esteem**. We used a subset of the Rosenberg Self-Esteem Scale [94] to measure individuals' global evaluation of the self. Although this scale is generally used to measure trait (or long-term) self-esteem [9], prior work suggested it can capture temporary changes in self-esteem after consuming content on Facebook for 3 minutes [43]. From the seven items used in a recent study on Facebook [31], we selected three items and modified them to refer to Instagram specifically.
- **Depression**. We employed the PHQ-2 to monitor depressive symptoms [6]. The PHQ-2 consists of the first two items of the PHQ-9 [61] and is widely used to screen for depression.

All scales were delivered in the local language: Korean. To ensure comprehension of the scales, we employed a double translation approach [76]. Two bilingual individuals participated independently in the translation process. The original English language was translated by a first translator into Korean and then was translated back to English by a second independent translator. The authors compared both versions and checked with the translators for inconsistencies.

3.3 Recruitment and Participants

Participants were recruited through word of mouth, social media, and the online forums of two universities, based in two different cities, in South Korea. Advertisements informed potential participants that the study took place over two weeks, that participation in the second week of the study was contingent on engagement levels during the first week and that, although it was about Instagram use on smartphones, it did not collect any posts or content from Instagram (e.g. news feed, direct message, etc). We screened for

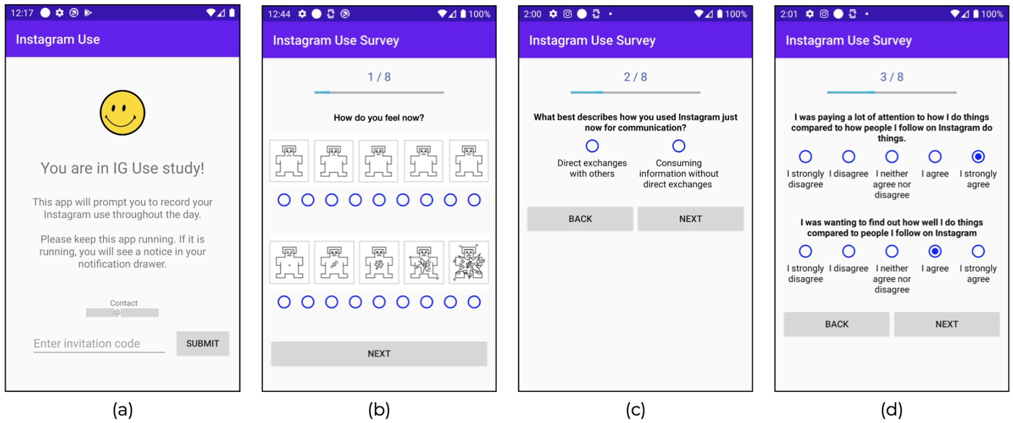


Figure 1: User interface of the IG Use study app: (a) home screen that displays the study protocol and (b), (c), and (d) showing example screenshots of the ESM survey. The screenshots were translated to English.

ownership of an Android smartphone and regular Instagram usage, defined as 30 minutes or more, on a daily basis, for the last six months [105]. We assessed Instagram usage levels by asking participants to share anonymized screen grabs of usage logs. Participants were compensated with the equivalent of $15 in local currency per week. Additionally, participants could receive a extra $5 by leaving the study app running for 80% or more of their screen time and a further $5 if they responded to over 50% of the survey completion notifications they received. These measures were designed to encourage participation and meant that, in total, study compensation ranged from between $15 and $40.

In total, nineteen participants were recruited. 16 participants were female and three male, with a mean age of 23.79 (SD=2.97). Three were recent graduates, five were graduate students and eleven were undergraduate students. 17 participants were South Korean and two were Chinese. During the study, all participants were based in South Korea. Using 5-point Likert scales, they self-reported a high familiarity with both computers (M=4.95, SD=0.23) and smartphones (M=4.47, SD=0.77) and were confident in their ability to understand Korean language conversation (M=4.79, SD=0.42). Logs indicate they had been Instagram users for between one and eight years (M=3.63, SD=2.03), had made between 2 and 566 posts (M=79.32, SD=143.05) in total and followed 86 to 694 other accounts (M=233.58, SD=174.68). During study enrollment, they self-reported using the service for between 30 minutes and 2 hours per day (M=58.11 minutes, SD=0.34), an above-average intensity of Instagram use (M=3.26, SD=1.03) [105] and a range of results on the PHQ-9 depression severity scale. Specifically, 11 were assessed as none to minimal, 5 were mild, 2 were moderate, and 1 was moderately severe. Participants reporting moderate severity and above were advised to consult an expert at their institution's healthcare center. One participant in the study were terminated due to a lack of engagement. Data from the remaining 18 are reported in the rest of this paper.

3.4 Procedure

Figure 2 displays an overview of the ESM study procedures. After screening, we first conducted an online orientation session to collect demographics, baseline questionnaire data (including PHQ-9 [61] and an adapted assessment of Instagram use [105]), consent, introduce the study protocol and walk through the study app installation process. Beyond installing the software, this involved setting permissions to allow the study app to monitor device application usage and adjusting battery optimization settings to prevent the study app from being placed in a sleep mode. Participants also used the software (both Instagram and survey interfaces) and were encouraged to ask questions or for clarifications of any points of confusion. We streamed the orientation session using Zoom. The orientation sessions lasted approximately 30 minutes and we held a total of six of these sessions: participants were given the choice of attending any of these at a time of their convenience. No more than five participants attended any single session. Scripts and slides were used to ensure that the same content was presented in all sessions. After a participant completed an orientation session, the study began immediately—participants began using Instagram while monitored by the study app and completing our ESM survey after some of these use sessions. During each participant's two week study period, we tracked their activity in real time. If no data was received for 48 hours, and we had received no notice a participant had withdrawn, we personally contacted participants and asked them to restart the study app. We also examined participants' results at the end of the first week to screen out participants who were not engaging with the study tasks. Specifically, if participants disabled our app for the majority of their screen time (50% or more) or answered the ESM survey less than once per day, we discontinued their participation in the study.

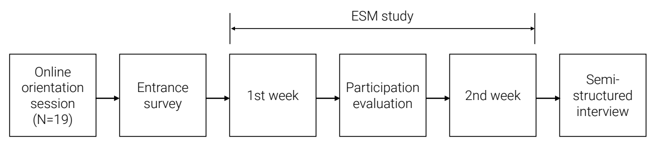


Figure 2: Procedure for the experience sampling study.

Table 1: Percentage of negative samples captured on each scale calculated by dividing data according to the standard practices for binary division. Cronbach's alpha is reported for constructs assessed by multiple items. Peak data are highlighted in bold.

Scale	Range	Cut point	Negative samples All *(N=837)*	Passive use *(N=633)*	Active use *(N=204)*	Cronbach's alpha
Valence	1-9	<5.0	12.54%	12.48%	**12.75%**	-
Arousal	1-9	>5.0	7.89%	6.32%	**12.75%**	-
Social comparison	1-5	>3.0	20.43%	**22.27%**	14.71%	0.27
Appearance comparison	1-5	>3.0	14.46%	**16.59%**	7.84%	0.86
Envy	1-5	>3.0	13.98%	**17.06%**	4.41%	0.92
Self-esteem	1-4	<2.5	8.60%	**10.43%**	2.94%	0.81
Depression	0-6	>= 3.0	12.66%	**14.85%**	5.88%	0.90

3.5 Results and Discussion

During the 14-day ESM study, participants used Instagram a total of 6878 times, corresponding to a mean of 342.74 times per person (SD=250.40) and 20.97 times per day (SD=14.26). This occupied a mean of 56.27 minutes (SD=30.79) per day in sessions that were a mean of 2.79 minutes (SD=6.79) long. In total, we notified participants to complete surveys after 2471 (35.93%) of these sessions (M=130.06, SD=65.96) and they responded to 37.88% of these prompts (887), ultimately completing 837 surveys (94.36% completion rate). 75.63% of the completed surveys were self-reported as involving passive use and the remaining 24.37% featured active use. On average, each survey took 32.23 seconds (SD=25.92) to complete. Overall, participants reported using Instagram frequently and in response to daily events and routines: before going to bed, while eating alone, or while traveling. They reported their activities on Instagram included watching Instagram stories, searching for restaurants with hashtags and checking notifications or viewing advertisements. They generally used Instagram to follow updates from friends or celebrities, watch the news, or share personal stories. Reflecting the high proportion of passive use sessions, four participants also noted they infrequently create posts, rather they consume content on the service.

In terms of affective experiences, we first sought to determine the most salient affective experiences reported in the survey data. We achieved this by examining the frequency with which different experiences were reported. This was calculated, for all scales except PHQ-2, by splitting the results at scale midpoint to create binary categories representing positive and negative experiences [107]. For PHQ-2, we followed standard procedures [62], and summed the number of responses that were three or greater to screen for depression. Table 1 depicts the prevalence of negative experiences across the constructs we studied. It displays the range, cut points and proportions of percentage of negative responses recorded by use type (active, passive or all). In addition, we calculated Cronbach's alpha score for each scale. The Cronbach's alpha score revealed inconsistencies in the items in the social comparison survey ($\alpha = 0.27$), so we dropped the social comparison scale from further analysis. However, given the literature indicating that social comparison is prevalent on Instagram in general [73], and within the Korean user population specifically [52], we attribute the inconsistencies we observed due to problems with the specific questionnaire items used; future work in this area should revisit the social comparison construct.

Despite these issues, the study results reveal the ubiquitous nature of both positive and negative experiences in SNS [33, 78, 88]. While prior researchers have studied the relationship between SNS usage and affective experiences (e.g., social comparison [34] or envy [114]), we study and report their prevalence during real world SNS experiences. This complements existing data—insights into the frequency with which particular affective states occurs can held ground judgements of how serious or important previously observed relationships actually are. Our data confirmed prior work suggesting that SNS use yields, in general, positive affective experiences [86]. Specifically, we found that 87.06% of all self-reports detailed neutral to positive experiences. Interview data also supported this claim. Participants reported generally positive outcomes after Instagram sessions. They viewed Instagram as a tool for relaxation, killing time, or resting. In a representative quote, P10 noted *"I think I generally feel better [after using Instagram], but I think I feel a little bad when I see myself using it for too long"*.

Behind the bulk of these generally positive experiences, however, are a range of more negative experiences. Perhaps unsurprisingly, these occur at quite different rates for different constructs. In our data, the prevalence rate of negative experiences during passive SNS use, shown in Table 1, peaks at 17.06% for envy and 16.59% for appearance comparison. Qualitative comments back up these spikes in the data. In our interviews, thirteen participants remarked on specific situations or occasions in which they experienced these negative emotions. Jealousy, evoked by images of events or activities that participants viewed as desirable, was a common theme. P2 summarized this as *"[..] When I saw people enjoying themselves outside, and when I saw pretty and cool influencers, I felt envious of them"*. Explicit comparisons, which could lead to negative emotions, were also reported. Some related to the image-heavy nature of Instagram, which P11 expressed as *"I'm following a lot of people who work out on Instagram and they post a lot of body pictures. [..] I think I am making a lot of comparisons with them"*. These results are largely confirmatory. Prior work has argued that SNS users experience relatively high levels of social comparison [52] and envy [108, 114] and engage in appearance comparison while browsing [34, 55]. Additionally, prior work has suggested image-centric social media platforms (e.g., Instagram) may be associated with body dissatisfaction [44], an effect that may be particularly prevalent in our predominantly young, female participant group [22]. Our findings add to the body of evidence for the presence of these ties.

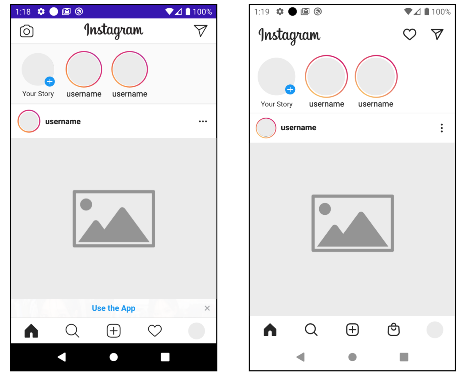
While not as prevalent, our users also reported a range of other negative emotions. However, these were often associated with forms of social comparison or envy. P6 expressed this well: *"When I hear stories of people who are in the same university but have achieved more things than me [..], I think my self-esteem is falling"*. This quote suggests that, although we can record moments of low self-esteem in our data, they may in fact be closely related to more frequently occurring feelings of envy or acts of comparison. As such, although prior work has identified self-esteem or depression as key affective concepts to study in the context of SNS use [3, 75], our results do not fully support their relevance. While self-esteem and depression can clearly vary during real world SNS use sessions, these variations may be due to a wide range of factors, including as consequences of the occurrence of more specific experiences such as envy or appearance comparison. In sum, while our results made it clear that the Instagram users we surveyed, in general, felt that their use of the service improved their affective state by relaxing, calming or entertaining them, they were also aware that specific types and forms of content prompted a range of more negative feelings. They expressed concern about the potential impact of these experiences on their mental well-being. Both our quantitative and qualitative data indicate that many of these more negative experiences took the form of either appearance comparison or envy. We therefore suggest that these constructs are particularly salient to Instagram use and deserve further study.

4 STUDY 2: DETECTING AFFECTIVE EXPERIENCES ON SNS

This study sought to determine whether (and how) smartphone sensor data can be used to detect affective experiences during real world SNS use. To achieve this, we designed a 14-day ESM study that logged sensor data while participants browsed their own Instagram feeds. ESM surveys, captured immediately after Instagram sessions, provided a self-assessment of affective experiences that we used as ground truth labels to train affect detection classifiers. This study was approved by the university's IRB.

4.1 Design and Materials

This study use broadly similar ESM methods to first: it again sought to capture experiences during genuine day-to-day social media use sessions. However, in addition to logging survey data about Instagram use, we also recorded smartphone sensor data during Instagram sessions. To achieve this, we redesigned *IG Use* as a foreground app that integrated sensor data logging functionality (see section 4.4.1) with an Android WebView [26] to display the Instagram service. We note IG Use did not record any contents displayed by Instagram. This approach enabled key features such as viewing news feed contents and stories, exploring content, viewing profiles, responding to content (e.g., liking, commenting), and sending direct messages. However, while these browsing and communicative activities are well supported by an embedded Instagram webview, general posting functionality is not available, and there are also some restrictions on viewing contents. Specifically, the native Instagram mobile application features several mechanisms



(a) (b)

Figure 3: Content feeds displayed on (a) IG Use and (b) Instagram. Presentation of contents, and browsing experience, is highly similar between the two systems. However, IG Use's implementation as an Android WebView does not enable it to support posting content of any form—the '+' button (center, bottom) typically used to upload contents is non-responsive. This is a limitation of the web view provided by the Instagram service.

for uploading content: creating a *post*, *story*, *reel* or *live stream*. Each of these options is associated with a different display format (e.g., in the feed, as the profile picture), upload mechanism (e.g., live or from existing files), display duration (e.g., permanent, or for 24 hours) and specific media type (e.g., image(s) and/or video(s) of varying lengths). In IG Use's approach of viewing Instagram in an embedded WebView, these options are unavailable—while the '+' icon typically used to access upload functionality is depicted, it does not provide access to any upload functionality. In addition, Instagram's embedded webview does not support display of *reels*: a recently introduced post format based around 15 second videos reminiscent of those shared via the TikTok SNS [77]. The updated IG Use app is open source and available for use, modification and download[1]. It is shown in Figure 3. We note that, excepting the posting and reel viewing features that were not available, IG Use provided a highly similar look and feel, and user experience during browsing, to the native Instagram application.

To assess this app's viability for the study purposes, we ran a short pilot in which six participants installed and used IG Use for one week. They reported it provided a satisfactory experience while browsing Instagram, validating its suitable for use in a full study. We also modified the ESM protocol in two key ways compared to study 1. Firstly, we limited the surveys to the constructs of appearance comparison and envy, as these previously showed high salience, and communication type, in order to retain our ability to distinguish between active and passive SNS use. Second, we adjusted the notification interval (the minimum gap between survey issuance) to 180 minutes to reduce the burden on, and frustration experienced by, our participants.

[1]https://github.com/mintra-ruensuk/IG-Use

4.2 Participants

Participants were recruited via online communities and word-of-mouth from three different academic institutions in three different cities in South Korea via advertisements for participation in a study on smartphone Instagram use that involved installing an app, letting it record sensor data in the background and completing multiple surveys per day. In order to ameliorate potential privacy concerns, the advertisements explicitly informed potential participants that the installed app did not record viewed content and thus no data about the items in their Instagram feeds would be logged; only surveys and phone sensor data [96]. Eligibility criteria were also maintained from the first study: ownership of an Android device and active use of Instagram (30 minutes or more per day) for at least 6 months. Financial incentives and engagement assessment processes were the same as the first study.

Twenty-two new participants joined this study (mean age 23.87, SD=3.09, 12 male and 10 female), all students (16 undergrad, five grad and one recent graduate) attending one of three different South Korean universities. All participants were South Korean and living in South Korea during the study. Using 5-point Likert scales, they self-reported a very high familiarity with both computers (M=4.93, SD=0.25) and smartphones (M=4.46, SD=0.74) and were highly confident of their ability to understand Korean language conversation (M=4.93, SD=0.25). They had held Instagram accounts for between 1 and 8 years (M=3.92, SD=2.37), created between 0 and 597 posts (M=99.4, SD=160.12), followed between 24 and 2174 other accounts (M=432.67, SD=537.22) and were followed by between 4 and 2174 accounts (M=432.67, SD=537.22). They self-reported using Instagram for 30 minutes to 6 hours per day (M=1.3 hour, SD=1.38) with an above-average intensity of Instagram use (M=3.22/5.0, SD=0.47). PHQ-9 results revealed their level of depression severity [61]: 13 were none to minimal, 6 were mild, 2 were moderate, and 1 was moderately severe. We again advised participants who scored over the cut point to consult an expert at their institution's healthcare center.

4.3 Procedure

Procedures, by and large, followed the first study. We first conducted a series of online orientation sessions with potential participants who responded to advertisements. In these 30 minute video conferences, participants were informed about the study's aim, protocol, data collection processes, and asked to provide consent. Those that agreed to do so were supported through the process of installing the IG Use app (as a sideloaded APK file), setting usage-tracking permissions and configuring battery optimization. They then filled in an entrance survey (identical to first study) and we explained about the Instagram functionality available in IG Use. Participants then used the app, including both to browse Instagram and to complete a survey. We specifically encouraged participants to use IG Use for browsing and the original Instagram app to create posts or stories (as this functionality was not available in IG Use) during the study period. Finally, we reminded participants they were free to terminate the study at any time. We again used slides and a script to ensure similar coverage of content in every orientation session and encouraged participants to ask questions at any time. For each participant, the study began immediately after completing an orientation session. After the study finished, participants completed an exit survey asking about the usability of, and their experiences with, IG Use, for examples of affective content on Instagram and general feedback about the study. They were then supported through the process of deleting the IG Use app.

4.4 Data Collection and Preprocessing

4.4.1 Features. We logged two types of features during IG Use sessions: 1) motion data from Motion sensors [27] and touch data from MotionEvents [28]. We selected these sensor channels for several reasons. First and foremost, prior work has reported that both motion data and touch data captured from mobile smart devices can be used to achieve accurate and reliable classification of affective, cognitive or behavioral states [39, 79, 96]. We sought to validate the promising performance reported in this prior work in the novel scenario of a field study of genuine social media use. In addition, such sensor systems are almost universally deployed on end-user devices: the overwhelming majority of modern smartphones feature both high resolution touch screens and high performance integrated Inertial Motion Units (IMUs) that track device movements. As our field study protocol involves deployment to participant's own phones we could therefore reasonably expect that use of these sensor systems would not present a barrier to participant recruitment. Furthermore, monitoring these sensor systems can be done without disrupting users (i.e., as a background process) and in a very wide range of environmental contexts such as noisy and/or dark environments. Finally, logging motion and touch data consumes limited device resources (e.g., battery power, processor time), requires relatively modest amounts of storage and has greatly reduced privacy implications compared to alternatives such as sensor channels based on image/video [81] or audio recording [16]. These factors further increase the practicality of touch and motion channels for deployment in a field study running on participants' own devices. In light of this rationale, we defined and logged data from 26 different motion and touch sensor channels, each of which is described in Table 2. All continuous features were logged at 60Hz.

4.4.2 Data Preprocessing. We analyzed the data using Scikit-learn [87], modelling our process after prior work in this area [96]. First, we catalogued the extent of missing or undefined data—1.8% of samples in total. Missing data was sporadic across participants and sensor channels and likely due to a very wide range of technical, power, resource conflict and network failures. As the proportion of missing data was relatively low, we opted to restore it via imputation. Specifically, we used a k-Nearest Neighbors approach based on the full set of unlabeled column values—all data bar the self-report items recording affective states. We then divided the data into non-overlapping windows of 15 seconds [119]. For each window, we calculated features, in the form of summary statistics (minimum, maximum, mean, median, standard deviation, and variance), for each sensor channel. We then conducted a normalization procedure to reduce the impact of features with difference ranges of magnitude and natural magnitude variations over time. Specifically, we employed two different normalization procedures. *Within-session* normalization was a standard normalization over all data within a

Table 2: Feature groups and specific features for each group.

Group	Features	Description
Motion	Acceleration (x, y, z)	Acceleration of the device along the three sensor axes.
	Gravity (x, y, z)	Direction and magnitude of gravity in the device's coordinates.
	Gyroscope (x, y, z)	Rate of rotation of the device around the three sensor axes.
	Light	Current illumination in SI lux units.
	Linear acceleration (x, y, z)	Linear acceleration of the device in the sensor frame, not including gravity.
	Magnetic field (x, y, z)	Ambient magnetic field, as measured along the three sensor axes.
	Rotation vector (x, y, z)	Orientation of the device relative to the East-North-Up coordinate frame obtained by integration of accelerometer, gyroscope, and magnetometer readings.
Touch	Distance	The overall distance of each touch in a 15s window.
	Hold duration	The duration of each touch in a 15s window.
	Inter-tap interval	The time between each touch in a 15s window.
	Touch area	The radius of each touch (sampled at 60Hz).
	Touch count	The number of times the screen was touched in 15s window.
	Touch pressure	The force applied during each touch (sampled at 60Hz).
	Speed	The overall movement speed of each touch in a 15s window.

given SNS use session. We also defined an alternative *prior-window* approach. This was only used on SNS use sessions that exceeded three minutes in length, or 26.06% of those we recorded. In these more prolonged sessions, we normalized data in each window with data from the preceding window. The intuition here was that affective states may vary during a session and thus normalization that adapted to changing baselines may be more effective.

To define labels to each window, we discretized the raw appearance comparison and envy values from the self-assessment surveys into binary classes using two different data binning approaches. First, we split data at scale mid-points [96], treating data at the mid-point as low-intensity. While intuitive, this approach yielded highly unequal class sizes—the occurrence rate of high scores on the appearance comparison (10.16%) and envy (16.37%) scales was low. Second, we applied equal-frequency binning [7], an approach which yields more equal class sizes.

We then constructed classifiers using this data and labels. Following prior work [7], we employed a leave-one-subject-out cross-validation procedure with a range of classifiers: ZeroR, AdaBoost (AB), Decision Tree (DT), k-Nearest Neighbour (kNN), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and RBF-kernel Support Vector Machine (SVM). We included the ZeroR classifier as a baseline: it simply predicts the majority class. As our study has 16 participants, we thus constructed 16 independent models and we report the mean performance over this set of models. For each model, we executed feature selection procedures and handled class imbalances in the data set. We opted not to conduct hyper-parameter tuning and simply used default settings for each algorithm in order to reduce the potential for over fitting. Feature selection steps were comprised of filter and wrapper methods [113]. In this process, we first removed constant features with zero variance and quasi-constant features with variance less than 1%. We then calculated the correlation matrix of the remaining features. If two or more features were highly correlated (Pearson's r greater than 0.8), we retained the one most highly correlated with the emotion labels. Next, we conducted recursive feature elimination using a Linear-kernel SVM with five-fold cross-validation on the remaining features. To handle the imbalanced data set, we over-sampled the minority classes (both high appearance comparison and high envy) using the Synthetic Minority Oversampling Technique (SMOTE) [17]. The feature selection and class balancing processes were only performed on the training data sets.

4.5 Results

4.5.1 Descriptive Statistics. Over the two-week study period, participants used IG Use 3860 times: an average of 175 times per person (SD=173.84), or 11.25 times per day. The mean IG Use session length was 3.77 minutes (SD=3.85). They used the original Instagram app an additional 6255 times, or 25.56 times a day (SD=20.96), spending a mean of 2.55 minutes (SD=2.18) each time. While this indicates that all of participants' SNS needs were not met by IG Use, adoption rates were sufficient to support data collection according to the ESM protocol. Indeed, we sampled surveys from participants during only a subset of IG Use sessions: we delivered 1222 notifications (to 31.66% of IG Use sessions) and participants responded to 869 of these (71.11%) or a mean of 72.20% per participant (SD=22.98). As expected, surveys indicated that activities on IG Use were predominantly passive (77.79%) rather than active (22.21%). Prior to processing the data further, we removed six participants due to low participation rates, ultimately leaving 837 self-report sessions and associated sensor data. Table 3 shows the distribution of binary affective classes in this data by construct (appearance comparison/envy) and binning technique (mid-point/equal frequency). As expected, mid-point binning led to class imbalances: participants tended to report low appearance comparison (89.84%) and low envy (83.63%). In terms of sensor data, we logged a large set of motion readings (30.22 million samples) and 69254 touches, which we divided into 7763 non-overlapping 15-second windows. This corresponds to 32 hours, 20 minutes, and 45 seconds of data.

Table 3: Descriptive statistics of self-reported scores, criteria for data binning, and the number of examples for each binning category. Values in square brackets indicate min and max scale scores in each range. *Appear.* refers to appearance comparison

		Criteria		# of self-reports		# of windows	
Data binning	**Label**	**Appear.**	**Envy**	**Appear.**	**Envy**	**Appear.**	**Envy**
Mid-point	Low (*majority class*)	<= 3	>3	752	700	6771	6438
	High (*minority class*)	<= 3	>3	85	137	992	1325
Equal-frequency	Low	[1.00,2.00]	[2.00,5.00]	480	357	3882	3881
	High	[1.00,2.33]	[2.33,5.00]	450	387	3881	3882

Table 4: Mean accuracy, class-wise F_1-scores (i.e., for both high (majority) and low (minority) classes), ROC AUC and MCC for binary appearance comparison and binary envy detection for all eight classifiers. Data in each window was normalized using our prior-window procedure and includes features from all sensors. Target labels were defined by using mid-point data binning.

Affective Dimension	Metric	ZeroR	AB	DT	kNN	LR	NB	RF	SVM
Appearance Comparison	Accuracy (%)	87.22	89.20	84.43	89.88	88.64	83.11	85.91	**90.98**
	Low (majority) F_1 (%)	93.17	91.01	88.47	90.62	85.41	90.84	89.59	**95.59**
	High (minority) F_1 (%)	00.00	50.89	54.89	64.69	43.73	29.44	56.05	**75.10**
	ROC AUC	0.50	0.87	0.84	0.85	0.80	0.71	0.86	**0.89**
	MCC	0.00	0.47	0.49	0.55	0.34	0.24	0.51	**0.70**
Envy	Accuracy (%)	82.93	89.67	77.79	89.25	88.19	85.66	88.23	**91.37**
	Low (Majority) F_1 (%)	90.67	91.23	85.79	91.35	92.36	89.85	92.59	**94.94**
	High (Minority) F_1 (%)	00.00	66.88	59.84	70.81	59.42	53.94	66.69	**71.87**
	ROC AUC	0.50	0.92	0.88	0.87	0.88	0.82	0.91	**0.92**
	MCC	0.00	0.60	0.57	0.65	0.49	0.35	0.64	**0.68**

4.5.2 Classifier Selection. We defined a *default* classifier configuration as follows: the use of data from both sensor channels, prior-window normalization and mid-point binning. We used this configuration to calculate the mean accuracy, the class-wise F_1-scores, the Area Under the Receiver Operating Characteristic Curve (ROC AUC) and the Matthews Correlation Coefficient (MCC) figures for the task of recognizing binary appearance comparison and envy for each of the eight classifiers. Table 4 shows the results. The headline outcome is that the RBF-kernel SVM outperformed other classifiers across the board. Accordingly, all subsequent analysis used this classifier. However, given the class imbalances in our data set, it is also worth commenting in detail on performance in the critical task of detecting the minority classes of high envy and high appearance comparison. Both ROC AUC and MCC data, metrics which are reported to be relatively immune to class imbalances [19], show good results with SVM. MCC scores, in particular, clearly differentiate between the effective performance achieved by the SVM model and that of other classifiers with broadly comparable scores on metrics such as accuracy (e.g., linear regression). While explicit interpretations of MCC scores should be treated with caution, we note the 0.68 and 0.7 scores attained by the SVM model are associated with strong to very strong levels of performance. The class-wise F_1-scores unpack this result. While they illustrate that performance in detecting the majority classes exceeded that of the minority classes by, respectively, of 20.49% and 23.07%, figures for the minority classes of high envy and high appearance comparison remained reasonable at 71.87% and 75.10%. We believe this represents a promising level of performance that is sufficient to support further investigation.

4.5.3 Classifier Variants. To provide a more nuanced picture of which aspects of the study contributed to classifier performance, we explored the impact of three variables: *the sensing modalities* used (all/motion/touch), the *normalization procedures* applied (within-session/prior-window), and the *data binning technique* (mid-point/equal-frequency) employed. Table 5 displays the accuracies, ROC AUC and MCC scores achieved using the RBF-kernel SVM recognizer with all 24 possible combinations of these variables. Class-wise F_1-scores for these configurations are in Table 6. The results indicate mixed performance profiles. In general, using motion data alone yielded improved accuracy (by up to 6.45%), ROC AUC (by up to 0.12) and MCC (by up to 0.15) in the majority of classifier configurations. In contrast, using touch data alone achieved strongly reduced performance. Based on these results, we suggest that motion data should be given precedence in mobile phone affect detection systems and the value of combining it with touch data should be carefully assessed—more data may not correspond to improved performance. A caveat to this claim is that the relatively weak performance of touch features in this study may have been due to the fact that the majority of use sessions we recorded were passive—they involved consumption rather than production of content. As such, they likely involved quite limited touch screen use. Other tasks, with a greater focus on user input, may yield different results.

A similar benefit may be present for prior-window normalization procedures: this generally boosts accuracy, ROC AUC and MCC by figures of as much as, respectively, 12.73%, 0.13 and 0.08 compared to our within-session approach. However, these effects

Table 5: Mean accuracy, ROC AUC and MCC for RBF-kernel SVM in predicting appearance comparison and envy using feature data from different *sensors*, different *normalization procedures*, and different *data binning* techniques.

	Affective Dim.	**Appearance comparison**				**Envy**			
	Data binning	**Mid-point**		**Equal-frequency**		**Mid-point**		**Equal-frequency**	
	Normalization	**Within**	**Prior**	**Within**	**Prior**	**Within**	**Prior**	**Within**	**Prior**
All features	Accuracy (%)	89.85	90.98	81.24	82.46	86.01	91.37	82.80	86.91
	ROC AUC	0.88	0.89	0.86	0.91	0.85	0.92	0.80	0.93
	MCC	0.65	0.70	0.61	0.68	0.63	0.68	0.65	0.73
Motion features	Accuracy (%)	95.78	91.76	79.99	82.87	92.46	93.95	87.48	89.83
	ROC AUC	0.93	0.91	0.81	0.89	0.91	0.96	0.92	0.94
	MCC	0.80	0.66	0.73	0.68	0.68	0.74	0.77	0.82
Touch features	Accuracy (%)	77.80	86.47	64.36	63.93	71.44	84.17	70.97	68.61
	ROC AUC	0.75	0.74	0.70	0.70	0.68	0.74	0.76	0.74
	MCC	0.32	0.29	0.34	0.32	0.31	0.29	0.46	0.44

Table 6: Class-wise classification F_1-score (in %) for RBF-kernel SVM in predicting appearance comparison and envy using feature data from different *sensors*, different *normalization procedures*, and different *data binning* techniques. We note that for mid-point binning high appearance comparison and high envy are minority classes, while for equal frequency binning, class sizes are purposely balanced: hence we label the different classes as High/Low throughout the table.

Affective Dim.	**Appearance comparison**								**Envy**							
Data Binning	**Mid-point**				**Equal-frequency**				**Mid-point**				**Equal-frequency**			
Normalization	**Within**		**Prior**		**Within**		**Prior**		**Within**		**Prior**		**Within**		**Prior**	
Class	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
All Features	94.1	64.4	95.6	75.1	80.0	78.3	81.9	78.2	91.0	67.2	94.9	71.9	83.4	82.8	86.3	83.4
Motion Features	97.7	78.8	90.4	70.5	81.4	78.3	84.8	80.3	92.5	80.0	96.4	81.1	86.1	88.6	90.8	88.5
Touch Features	86.2	43.7	71.4	47.1	64.1	64.6	64.6	63.2	81.2	40.7	91.1	30.1	67.7	73.6	70.5	66.4

were not universal. While our prior-window technique offers advantages in many cases, peak accuracy for appearance comparison was achieved using our within-session approach. We conclude that larger and more diverse data sets will likely be required to make comprehensive recommendations on best practices for data normalization.

Results for binning technique were clearer. Mid-point binning, which creates class imbalances but divides the self-report data at a more meaningful threshold, led to consistently and substantially improved accuracies for both affective dimensions (of up to 22.54%), while also achieving strong results in minority class F_1-scores: up to 78.8% and 81.1% for, respectively, appearance comparison and envy. While ROC AUC and MCC data remained consistently high for both binning techniques, these metrics are more compelling for the imbalanced classes of mid-point binning than the balanced ones created using the equal-frequency approach (a situation in which the accuracy metric typically performs well). We suggest these results indicate that equal-frequency binning blurred the boundaries between the actual affective states being assessed, reducing the quality of classification that could be achieved. As such, future work should carefully consider the pros and cons of creating balanced classes over those of dealing with imbalanced classes, split at more inherently meaningful thresholds. Our results suggest approaches such as equal-frequency binning may be best avoided.

To further explore this assertion, we split the test sets for each of the equal-frequency LOOCV models into two halves: one containing data from windows labelled with affective scores *near* to the binning thresholds and one containing data from windows labelled with scores *far* from the binning thresholds. This corresponds to data in which, respectively, less extreme and more extreme affective states were reported. We achieved this simply by splitting the data into four equally sized bins based on the affective ratings, then combining data from the two central bins (second and third quartiles) to create a *near* test subset and the two extreme bins (first and fourth quartiles) to create a *far* test subset. We report on the performance achieved using these two test subsets in Table 7. The results from this analysis provide further evidence to support our assertions about the inappropriateness of the equal-frequency binning approach: mean performance with the *far* test subsets outperformed that with the *near* subsets modestly, but consistently, across all metrics. This suggests that the affective states experienced in the *near* test subsets were less clearly differentiated, in terms of the observable behaviors they generated, than those in the *far* test subsets. Based on these results, we conclude that, at least for the affective constructs we studied in this work, and specific scales we used, equal-frequency binning did not support a division of data into optimally separably affective classes. We suggest that future work will need explore alternative methods to deal with the class imbalances that will inevitably be present in field studies of naturally occurring affect.

4.5.4 Feature Performance. To explore the specific features that contributed to recognizer performance, we used the *default* classifier configuration to generate representative sets of 28 features for both appearance comparison and envy by conducting the feature selection process on the full set of prior-window normalized data in

Table 7: Mean accuracy, class-wise F_1-scores (i.e., for both high and low affect classes), ROC AUC and MCC for binary appearance comparison and binary envy detection. Results based on the SVM classifier, both within-session and prior-window normalization, equal-frequency binning and data from all sensors. In addition, test sets were divided into two equally sized subsets involving affective rating scores *near* to (second and third quartile) and *far* from (first and fourth quartile) the binning threshold.

Affective Dim.	Appearance comparison				Envy			
Normalization	Within		Prior		Within		Prior	
Test set	Near	Far	Near	Far	Near	Far	Near	Far
Accuracy	80.23	82.25	80.66	84.25	80.97	84.62	84.87	88.95
Low (F_1)	77.92	82.04	80.21	83.65	79.48	87.34	82.21	90.43
High (F_1)	75.73	80.91	75.86	80.46	79.06	86.62	83.33	83.41
ROC AUC	0.85	0.89	0.89	0.93	0.76	0.84	0.91	0.94
MCC	0.61	0.63	0.66	0.66	0.65	0.68	0.72	0.75

Table 8: The features showing the greatest predictive power for detecting binary appearance comparison and binary envy. Features shown in alphabetical order. *min*=minimum, *max*=maximum, *var*=variance, *std*=standard deviation

Feature	Sensor	Appearance	Envy
Acc., x-axis	Motion	min	max, mean
Acc., y-axis	Motion	max	min, mean
Acc., z-axis	Motion	min	-
Gravity, x-axis	Motion	min	max, mean
Gravity, y-axis	Motion	-	min, max
Gravity, z-axis	Motion	min	-
Gyroscope, x-axis	Motion	max	max, mean
Gyroscope, y-axis	Motion	max	min
Gyroscope, z-axis	Motion	min, max	-
Inter-tap interval	Touch	max, mean, var	mean, var
Light	Motion	max	-
Linear acc., x-axis	Motion	max, mean	var
Linear acc., y-axis	Motion	min, max, mean	min
Linear acc., z-axis	Motion	min	min
Magnetic field, x-axis	Motion	min	-
Magnetic field, y-axis	Motion	min, max	min, max, mean, var
Magnetic field, z-axis	Motion	max	mean
Rotation vector, x-axis	Motion	-	var
Rotation vector, z-axis	Motion	min	min, max
Touch pressure	Touch	min, max, mean, var	min, max, mean, std

the study. These are shown in Table 8. Motion features dominated contributing, respectively, 21 (75%) and 22 (78.57%) of features for appearance comparison and envy. This reinforces the conclusion that touch features had limited value in our study.

4.6 Discussion

The goal of this study was to explore whether affect detection systems for smartphones, a technology previously demonstrated in numerous studies involving lab evaluations [38, 51, 96] or performance of artificial repetitive tasks [79], can scale to a meaningful real world use scenario—day-to-day browsing one's own SNS feed. The results are, by and large, affirmative and validate prior work in this area. Our affective state influences how we operate or move our phones in ways that are sufficiently systematic to support accurate detect of affect. Moving beyond this positive conclusion, it is worth contrasting the results of the current study with prior work in detail. Most importantly, we note that classifier performance remained high in our study (93.95% to 95.78% peak accuracy). We argue this level of performance is not only sufficient to support the design of affect aware features or services, but also on par with the prior work that has conducted studies in more controlled settings. For example, in closely related work detecting binary valence and arousal during use of a different SNS (Facebook), Ruensuk et al. [96] report accuracies as high as 94.16%, roughly equivalent to the performance achieved here. Similar performance levels have been reported in a wide range of other controlled settings, such as 89.1% accuracy in binary classification of elicited affect [79] or an AUC of up to 0.84 in the task of classifying along three levels of valence [115].

While study differences render precise comparisons invalid, the strong performance we observe is somewhat unexpected, as data captured in natural settings could reasonably be assumed to noisy, obscured or otherwise distorted by issues as mundane as the characteristics of different device models [79]. We attribute these strong results to, in part, our selection of appropriate affective constructs—our first study indicated that the fine-grained notions of appearance comparison and envy are more relevant to day-to-day Instagram use than the broader constructs of valence and arousal. As such, we suggest that targeting specific, application relevant affective constructs may help support strong classification performance in affect detection systems. Appropriately granular constructs will not only allow a research or design team to focus on specific aspects of an affective experience, but may also increase the quality of the labelling used to train classifiers. For example, in our case, while valence and arousal may fluctuate during an ESM study period for any number of reasons (e.g., prevailing mood, time of day, external events) [29], constructs such as appearance comparison and envy assessed just after browsing are highly likely to relate to the viewed contents. In addition, we note there may simply be advantages to building affect detection systems based on genuine user experiences—affect that occurs in the wild, as opposed to that elicited in the lab—may be more distinctively and consistently expressed. While further field or ESM studies would be required to fully support these claims, we believe our current results are sufficiently persuasive to stimulate further real world studies in this area. We see a strong future for smartphone affect detection systems developed with user data from genuine lived experiences.

5 GENERAL DISCUSSION AND DESIGN IMPLICATIONS

This paper aimed to increase the granularity of the affective experiences captured during studies of real world SNS use and to investigate whether or not the day-to-day affective experiences of SNS users can be detected automatically using data from smartphone sensors. To achieve this, we first conducted a two-week ESM study that captured a wide range of finely grained affective constructs directly after an SNS use session. The results indicate that the most frequent negative affective experiences were appearance comparison and envy. Building on this finding, we conducted a second ESM study to collect smartphone sensor data alongside ESM self-reports of the affective experiences recorded immediately after the SNS browsing sessions. Classifiers built using the data and labels captured in this study were able to predict binary appearance comparison with an accuracy of up to 95.78% and binary envy with an accuracy of up to 93.95%. These results suggest it is feasible to detect specific, targeted negative affective experiences while consuming content on SNS using the sensors integrated into standard smartphones. We believe the high performance levels we achieve are sufficient to support design of diverse applications and services.

An important next step for this work will be to flesh out the design space for such affect aware services. This will be a challenging task—while literature in this area is in its infancy, initial studies paint a complex portrait of user reactions to this type of technology: an acknowledgement of the potential of such services to increase well-being or mitigate harm [93] combined with a widely prevalent vision of emotions as fundamentally private and interventions based on detecting them as risky tools that could lead to losses of autonomy [2]. We identify a clear need for further formative work (e.g., using methods such as Zimmerman and Forlizzi [121]'s speed dating) to increase our understanding of user needs for, and perceptions about, affect aware technology on SNS. Critical aspects that should be fore-fronted in this work include issues of privacy (i.e., in terms of what data is used to achieve affect awareness) and control (i.e., in terms of whether or not designs are perceived to empower users or other stakeholders such as the SNS itself).

While the work in this paper does not explicitly tackle these issues, key aspects of our system and studies do point to potentially desirable qualities for future designs in this space. Perhaps the most salient of these is granularity. The majority of SNS use sessions in our work did not result in undesirable affective experiences; negative affect was an infrequent outcome. This suggests that not all SNS use sessions are created equal and ties well with longstanding claims that SNSs cannot be effectively considered as monolithic entities: each includes diverse features which users deploy to meet various needs [104]. We argue that the notion that SNS use should be unpacked (or "unbundled") to be understood also applies to the experiences users have on these services: different SNS features, and types of content, may evoke different affective experiences and effective designs for affect aware services need to be sensitive to these variations - they need to be highly granular. Anecdotal evidence to support this claim comes from recent work (using similar contextually-triggered ESM methods) that examines regretful feature level smartphone use of SNSs [20]. This work discusses how feature-level problematic usage patterns could be conveyed to users to promote self-reflection and self-regulation and argues that such fine-grained feedback will be more effective than coarse techniques such as global limits on app use sessions or time.

We believe the design space for affect aware services that operate at a highly granular level is large. For example, we could design logs and visualizations that highlight how the prevalence of key affective states varies with use of different SNS features, providing users with the ability to extract novel insight about the implications of their specific SNS use behaviors. Prior work has suggested that increased understanding of the emotional consequences of behaviors can help support behavior change [50]. A more direct design in this space might immediately notify users when particular affective states (such as negative appearance comparison) are detected, highlighting the specific SNS features associated with the experience, and recommending (or enforcing) amelioratory actions such as a break or pause in browsing [59]. An even greater level of granularity, designs might associate different affective experiences with different classes of content, such as those posted from different sources (e.g., influencers, public figures, friends) or that contain different types of content (e.g., derived from image hash tags). Such systems could potentially recommend or curate content based on detected affect [89]. For example, by asking a user whether filters should be activated to display fewer feed contents that evoke negative affective experiences such as appearance comparison and envy. While the use of such filters may result in reduced control and autonomy, they may ultimately be valuable tools for specific vulnerable populations such as teens with pro-eating disorder who have developed negative feelings about their body image on SNSs [15].

Beyond these practical implications for the design of SNS, it is also worth discussing the implications of our work for the wider research community who might seek to apply related methods in other contexts, such as on different SNS or with different affective states. Our study designs were predicated on key two observations. First, broad academic evidence [22], and public perception [116], that Instagram is fertile ground for negative affective experiences in many users of the service and particularly among young adults and teenagers. Second, that Instagram use sessions tend to be both frequent and short—two to three minute sessions 20 to 30 times per day in the studies reported here. Leveraging these properties, our ESMs were able to capture a reasonable proportion of well-delimited use sessions that resulted in negative affect in a relatively short (two-week) study period. However, many other problematic SNS use patterns may not combine these properties: doomscrolling [12], for example, is associated with prolonged use, often in the period immediately prior to sleep, while addictive watching of videos on YouTube [71] also involves sustained viewing. While both these situations are associated with negative affect (respectively, anxiety and a loss of agency), the ESMs we deploy in this paper would not be directly applicable to these use contexts: while appropriate questionnaires could be included to address relevant affective constructs, delivering timely notifications and segmenting data into well-labelled chunks would be challenging. As such, we conclude that the methods we employ in this paper are mainly applicable to SNS use patterns involving sessions that are relatively brief or can be otherwise readily segmented. There would need to be substantial modification to the protocols for sampling self reports and for associating logged sensor with labelled affective states if our methods were applied to scenarios involving more prolonged sessions.

Finally, it is also worth commenting on some of the technical aspects of our second study and highlighting avenues for future work. We used sensor channels (e.g., touch, motion), statistical features (e.g., min, max, mean, etc.) and machine learning classifiers (e.g., SVM) that are both relatively simple and mature, with various prior demonstrations of their capability to realize affect detection systems on mobile devices [79, 115], albeit in more limited controlled and/or lab contexts. While this strong prior performance provides a clear rationale for these choices, and ensures that our relatively simple sensing and classification systems are readily executable on resource constrained mobile devices, we also note they leave considerable scope for technical innovation. Future work should examine emerging sensor channels, such as eye gaze [96] or physiological measurements such as heart rate [117], seek to generate new feature types that may support improved performance, such as image based depictions [115] and use these in conjunction with state of the art classifiers such as those based on deep learning approaches [95]. We believe such developments can help ensure that the promising performance we report can be maintained in broader populations and use contexts.

6 LIMITATIONS

There are a number of limitations that may impact the work reported in this paper. First and foremost, our ESM methods left gaps in our ability to capture SNS use. While continuous sampling is not the goal of an ESM study, these problems were particularly evident in the second study. Specifically, due to technical limitations, our study app did not implement Instagram functionality relating to creating posts and other forms of content. As such, participants completed the majority of their Instagram sessions in the original app. This not only represents missed data, but may also mean that the sessions in which participants used our app were not entirely natural—they may have explicitly chosen to use IG Use to complete study tasks, knowing they would not be able to freely use all Instagram features. This may have biased the data we were able to record. Future work in this area should carefully explore the pros (capture sensor data readily) and cons (incomplete implementation of SNS features that may lead to altered use patterns) of developing bespoke study apps using technologies such as Android WebView.

Other limitations relate to our sample: due in part to the specific screening procedures it is limited in size and, by design, targets relatively heavy SNS users. In addition, our participants were a relatively homogeneous group of east Asian university students based in South Korea and our first study, in particular, featured a large proportion of female participants, a demographic that has been previously associated with problematic Instagram use patterns [22]. As such, the results we report may not fully generalize to other populations; further work at larger scales would be required to increase confidence in the results we report and to comprehensively assess generalizability. On a positive note here, we believe that the experience sampling methods we use are inherently scalable, suggesting the barriers to conducting larger studies on more diverse populations will be practical (e.g., budgetary, logistical) rather than more fundamental. Indeed, to support such future efforts, we have open-sourced our IG Use app. Beyond the limitations of our sample, our results may also not generalize to other SNS platforms. Much of the data we report may be of most relevance to image-centered services that feature high levels of passive use, such as Instagram, and less pertinent to more general services such as those focused on diverse forms of media or status update (e.g., Facebook), those that specialize on other specific types of content, such as video media (e.g., TikTok [77]), or those have a higher prevalence of active use. In order to determine the viability of affect detection services for social media in general, future research should seek to catalog the key affective states users experience during use of different services then develop customized ESM procedures to appropriately capture and segment SNS use sessions. Given the diversity of services available, and the different use styles they support, we expect this work will be quite specific to the detail and nuances of each service.

Finally, it is worth commenting on the limitations inherent in our choices of the specific affective constructs we studied. As a general model of emotion in our first study, we focused on the circumplex model [97]: a two dimensional system with major axes of valance and arousal. In addition, we opted to discretize these constructs (and others we studied) in two binary classes. While these choices are common in the literature on affect detection in HCI [79, 96], they are far from universally applied. For example, in closely related work deploying field study methods for affect detection, Zhang et al. [118] opt to use Ekman's six basic emotions (happiness, sadness, anger, surprise, fear, disgust) [30] arguing that they are readily understandable by users and support the intuitive idea that multiple different emotions can simultaneously co-exist at different intensities. While our studies are not designed to make recommendations about the most appropriate general models for emotion to use in studies of affect detection, we do identify this as a key issue for future work: selecting appropriate affective models will increase the effectiveness and understandability of affect detection systems while selecting appropriate discretizations will optimize their resolution. The fact we ultimately retained a pair of specific affective constructs (binary appearance comparison and envy) for further study does support the assertion that discrete models, such as Ekman's [30], may be preferable. However, as literature also suggests that different models may be best suited to different individuals [4], we also note that customization may be the best approach. Future work should be designed to explore and shed light on these complex issues.

7 CONCLUSION

In conclusion, the work in this paper documents the prevalence of different affective constructs during real world social media use. It suggests that envy and appearance comparison are particularly relevant and deserve more attention in future work. In addition to documenting the frequency with which these affective states occur, we demonstrate that these critical moments can be detected with a high degree of accuracy (up to 95.78%) with the sensors on a standard smartphone. This work opens the door to the design of affect aware SNS services that can support their users as they seek to relax and enjoy social media by shielding them from content that provokes more negative affect. Designing services that do this by empowering and enabling their users, rather than simply restricting or concealing content from them, will be a challenging task. Exploring the design space for such services is the next step for this work.

ACKNOWLEDGMENTS

We extend our thanks to all our experimental participants and to Eunyong Cheon for his help in running the studies reported in this paper. This work was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) and funded by the Korean government (NRF-2021M3A9E4080780) and the Basic Science Research Program through the National Research Foundation of Korea funded by the Korean government (NRF-2020R1F1A1066408).

REFERENCES

[1] June Ahn and Jinyoung Kim. 2013. The Show Must Go on: The Presentation of Self During Interpersonal Conflict on Facebook. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries* (Montreal, Quebec, Canada) *(ASIST '13)*. American Society for Information Science, Silver Springs, MD, USA, Article 94, 10 pages. http://dl.acm.org/citation.cfm?id=2655780.2655874

[2] Nazanin Andalibi and Justin Buss. 2020. *The Human in Emotion Recognition on Social Media: Attitudes, Outcomes, Risks.* Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376680

[3] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2017. Sensitive Self-Disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1485–1500. https://doi.org/10.1145/2998181.2998243

[4] Lisa Feldman Barrett. 1998. Discrete Emotions or Dimensions? The Role of Valence Focus and Arousal Focus. *Cognition and Emotion* 12, 4 (1998), 579–599. https://doi.org/10.1080/026999398379574 arXiv:https://doi.org/10.1080/026999398379574

[5] Lisa Feldman Barrett and Daniel J Barrett. 2001. An introduction to computerized experience sampling in psychology. *Social Science Computer Review* 19, 2 (2001), 175–185.

[6] Till Beiwinkel, Stefan Hey, Olaf Bock, and Wulf Rössler. 2017. Supportive Mental Health Self-Monitoring among Smartphone Users with Psychological Distress: Protocol for a Fully Mobile Randomized Controlled Trial. *Frontiers in Public Health* 5 (2017), 249. https://doi.org/10.3389/fpubh.2017.00249

[7] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 221, 12 pages. https://doi.org/10.1145/3290605.3300451

[8] Jennifer L. Bevan, Jeanette Pfyl, and Brett Barclay. 2012. Negative emotional and cognitive responses to being unfriended on Facebook: An exploratory study. *Computers in Human Behavior* 28, 4 (2012), 1458–1464. https://doi.org/10.1016/j.chb.2012.03.008

[9] Roberta Biolcati and Stefano Passini. 2018. Narcissism and self-esteem: Different motivations for selfie posting behaviors. *Cogent Psychology* 5, 1 (2018), 1437012.

[10] danah m. boyd and Nicole B. Ellison. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 1 (10 2007), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x arXiv:https://academic.oup.com/jcmc/article-pdf/13/1/210/22316979/jjcmcom0210.pdf

[11] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (1994), 49 – 59. https://doi.org/10.1016/0005-7916(94)90063-9

[12] Kathryn Buchanan, Lara B Aknin, Shaaba Lotun, and Gillian M Sandstrom. 2021. Brief exposure to social media during the COVID-19 pandemic: Doom-scrolling has negative emotional consequences, but kindness-scrolling does not. *Plos one* 16, 10 (2021), e0257728.

[13] Moira Burke and Robert Kraut. 2013. Using Facebook after Losing a Job: Differential Benefits of Strong and Weak Ties. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1419–1430. https://doi.org/10.1145/2441776.2441936

[14] Pew Research Center. 2021. Social Media Fact Sheet. https://www.pewresearch.org/internet/fact-sheet/social-media/

[15] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 1201–1213. https://doi.org/10.1145/2818048.2819963

[16] Keng-hao Chang, Drew Fisher, John Canny, and Björn Hartmann. 2011. How's My Mood and Stress? An Efficient Speech Analysis Library for Unobtrusive Monitoring on Mobile Phones. In *Proceedings of the 6th International Conference on Body Area Networks* (Beijing, China) *(BodyNets '11)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 71–77.

[17] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[18] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1217–1230. https://doi.org/10.1145/2998181.2998213

[19] Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21, 1 (02 Jan 2020), 6. https://doi.org/10.1186/s12864-019-6413-7

[20] Hyunsung Cho, DaEun Choi, Donghwi Kim, Wan Ju Kang, Eun Kyoung Choe, and Sung-Ju Lee. 2021. Reflect, Not Regret: Understanding Regretful Smartphone Use with App Feature-Level Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 456 (oct 2021), 36 pages. https://doi.org/10.1145/3479600

[21] Tak Sang Chow and Hau Yan Wan. 2017. Is there any 'Facebook Depression'? Exploring the moderating roles of neuroticism, Facebook social comparison and envy. *Personality and Individual Differences* 119 (2017), 277–282. https://doi.org/10.1016/j.paid.2017.07.032

[22] Rachel Cohen, Toby Newton-John, and Amy Slater. 2017. The relationship between Facebook and Instagram appearance-focused activities and body image concerns in young women. *Body Image* 23 (2017), 183–187. https://doi.org/10.1016/j.bodyim.2017.10.002

[23] Ahuti Das-Friebel, Anita Lenneis, Anu Realo, Adam Sanborn, Nicole KY Tang, Dieter Wolke, Adrian von Mühlenen, and Sakari Lemola. 2020. Bedtime social media use, sleep, and affective wellbeing in young adults: an experience sampling study. *Journal of Child Psychology and Psychiatry* 61, 10 (2020), 1138–1149.

[24] Munmun De Choudhury, Scott Counts, and Michael Gamon. 2012. Not all moods are created equal! exploring human emotional states in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*. Association for the Advancement of Artificial Intelligence, Palo Alto, CA, USA, 8 pages.

[25] Dian A. de Vries, A. Marthe Möller, Marieke S. Wieringa, Anniek W. Eigenraam, and Kirsten Hamelink. 2018. Social Comparison as the Thief of Joy: Emotional Consequences of Viewing Strangers' Instagram Posts. *Media Psychology* 21, 2 (2018), 222–245. https://doi.org/10.1080/15213269.2016.1267647 arXiv:https://doi.org/10.1080/15213269.2016.1267647

[26] Android Developers Documentation. 2021. Building web apps in WebView. https://developer.android.com/guide/webapps/webview. Accessed: 2021-03-29.

[27] Android Developers Documentation. 2021. Motion sensors. https://developer.android.com/guide/topics/sensors/sensors_motion.

[28] Android Developers Documentation. 2021. MotionEvent. https://developer.android.com/reference/android/view/MotionEvent.

[29] Boris Egloff, Anja Tausch, Carl-Walter Kohlmann, and Heinz Walter Krohne. 1995. Relationships between time of day, day of the week, and positive mood: Exploring the role of the mood measure. *Motivation and emotion* 19, 2 (1995), 99–110.

[30] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* 6, 3-4 (1992), 169–200. https://doi.org/10.1080/02699939208411068 arXiv:https://doi.org/10.1080/02699939208411068

[31] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12, 4 (07 2007), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x arXiv:https://academic.oup.com/jcmc/article-pdf/12/4/1143/22316419/jjcmcom1143.pdf

[32] Facebook. 2020. Facebook Reports Fourth Quarter and Full Year 2020 Results. https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx

[33] Lien Faelens, Kristof Hoorelbeke, Bart Soenens, Kyle Van Gaeveren, Lieven De Marez, Rudi De Raedt, and Ernst H.W. Koster. 2021. Social media use and well-being: A prospective experience-sampling study. *Computers in Human Behavior* 114 (2021), 106510. https://doi.org/10.1016/j.chb.2020.106510

[34] Jasmine Fardouly and Lenny R. Vartanian. 2015. Negative comparisons about one's appearance mediate the relationship between Facebook usage and body image concerns. *Body Image* 12 (2015), 82–88. https://doi.org/10.1016/j.bodyim.2014.10.004

[35] Giulia Fioravanti, Alfonso Prostamo, and Silvia Casale. 2020. Taking a Short Break from Instagram: The Effects on Subjective Well-Being. *Cyberpsychology, Behavior, and Social Networking* 23, 2 (2020), 107–112.

[36] Joel E. Fischer and Steve Benford. 2009. *Inferring Player Engagement in a Pervasive Experience.* Association for Computing Machinery, New York, NY, USA, 1903–1906. https://doi.org/10.1145/1518701.1518993

[37] Jesse Fox and Jennifer J. Moreland. 2015. The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in Human Behavior* 45 (2015), 168 – 176. https://doi.org/10.1016/j.chb.2014.11.083

[38] Yuan Gao, Nadia Bianchi-Berthouze, and Hongying Meng. 2012. What Does Touch Tell Us About Emotions in Touchscreen-Based Gameplay? *ACM Trans. Comput.-Hum. Interact.* 19, 4, Article 31 (Dec. 2012), 30 pages. https://doi.org/10.1145/2395131.2395138

[39] Surjya Ghosh, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2017. TapSense: Combining Self-Report Patterns and Typing Characteristics for Smartphone

Based Emotion Detection. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) *(MobileHCI '17)*. Association for Computing Machinery, New York, NY, USA, Article 2, 12 pages. https://doi.org/10.1145/3098279.3098564

[40] Surjya Ghosh, Kaustubh Hiware, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2019. Emotion detection from touch interactions during text entry on smartphones. *Int. J. Hum.-Comput. Stud.* 130 (2019), 47–57.

[41] Frederick X Gibbons and Bram P Buunk. 1999. Individual differences in social comparison: development of a scale of social comparison orientation. *Journal of personality and social psychology* 76, 1 (1999), 129.

[42] Amy L. Gonzales. 2014. Text-based communication influences self-esteem more than face-to-face or cellphone communication. *Computers in Human Behavior* 39 (2014), 197–203. https://doi.org/10.1016/j.chb.2014.07.026

[43] Amy L Gonzales and Jeffrey T Hancock. 2011. Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, behavior, and social networking* 14, 1-2 (2011), 79–83.

[44] Scott Griffiths, Stuart B Murray, Isabel Krug, and Sian A McLean. 2018. The contribution of social media to body dissatisfaction, eating disorder symptoms, and anabolic steroid use among sexual minority men. *Cyberpsychology, Behavior, and Social Networking* 21, 3 (2018), 149–156.

[45] Scott Griffiths and Ashleigh Stefanovski. 2019. Thinspiration and fitspiration in everyday life: An experience sampling study. *Body Image* 30 (2019), 135–144. https://doi.org/10.1016/j.bodyim.2019.07.002

[46] Sharath Chandra Guntuku, Daniel Preotiuc-Pietro, Johannes C Eichstaedt, and Lyle H Ungar. 2019. What twitter profile and posted images reveal about depression and anxiety. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. Association for the Advancement of Artificial Intelligence, Palo Alto, CA, USA, 236–246.

[47] Maryam Hasan, Elke Rundensteiner, and Emmanuel Agu. 2014. Emotex: Detecting emotions in twitter messages. In *Proceedings of the ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference (ASE'14)*. Academy of Science and Engineering (ASE), USA, 27–31.

[48] Jan Crusius Helmut Appel and Alexander L. Gerlach. 2015. Social Comparison, Envy, and Depression on Facebook: A Study Looking at the Effects of High Comparison Standards on Depressed Individuals. *Journal of Social and Clinical Psychology* 34, 4 (2015), 277–289. https://doi.org/10.1521/jscp.2015.34.4.277 arXiv:https://doi.org/10.1521/jscp.2015.34.4.277

[49] Kristin E Heron and Joshua M Smyth. 2010. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British journal of health psychology* 15, 1 (2010), 1–39.

[50] Victoria Hollis, Artie Konrad, and Steve Whittaker. 2015. Change of Heart: Emotion Tracking to Promote Behavior Change. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2643–2652. https://doi.org/10.1145/2702123.2702196

[51] Sinh Huynh, Seungmin Kim, JeongGil Ko, Rajesh Krishna Balan, and Youngki Lee. 2018. EngageMon: Multi-Modal Engagement Sensing for Mobile Games. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 13 (March 2018), 27 pages. https://doi.org/10.1145/3191745

[52] Ha Sung Hwang. 2019. Why social comparison on Instagram matters: Its impact on depression. *KSII Transactions on Internet and Information Systems (TIIS)* 13, 3 (2019), 1626–1638.

[53] Nicholas C Jacobson, Hilary Weingarden, and Sabine Wilhelm. 2019. Using digital phenotyping to accurately detect depression severity. *The Journal of nervous and mental disease* 207, 10 (2019), 893–896.

[54] Shaohai Jiang and Annabel Ngien. 2020. The Effects of Instagram Use, Social Comparison, and Self-Esteem on Social Anxiety: A Survey Study in Singapore. *Social Media + Society* 6, 2 (2020), 2056305120912488. https://doi.org/10.1177/2056305120912488 arXiv:https://doi.org/10.1177/2056305120912488

[55] S Venus Jin and Ehri Ryu. 2020. "I'll buy what she's# wearing": The roles of envy toward and parasocial interaction with influencers in Instagram celebrity-based brand endorsement and social commerce. *Journal of Retailing and Consumer Services* 55 (2020), 102121.

[56] Caitlyn Johnston and William E. Davis. 2019. Motivating exercise through social media: Is a picture always worth a thousand words? *Psychology of Sport and Exercise* 41 (2019), 119–126. https://doi.org/10.1016/j.psychsport.2018.12.012

[57] Maria Kalpidou, Dan Costin, and Jessica Morris. 2011. The relationship between Facebook and the well-being of undergraduate college students. *CyberPsychology, behavior, and social networking* 14, 4 (2011), 183–189.

[58] Małgorzata Karpińska-Krakowiak and Artur Modliński. 2018. The Effects of Pranks in Social Media on Brands. *Journal of Computer Information Systems* 58, 3 (2018), 282–290. https://doi.org/10.1080/08874417.2016.1241683 arXiv:https://doi.org/10.1080/08874417.2016.1241683

[59] Jaejeung Kim, Hayoung Jung, Minsam Ko, and Uichin Lee. 2019. GoalKeeper: Exploring Interaction Lockout Mechanisms for Regulating Smartphone Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 16 (mar 2019), 29 pages. https://doi.org/10.1145/3314403

[60] Hsiu-Chia Ko and Dian-Han Yu. 2019. Understanding Continuance Intention to View Instagram Stories: A Perspective of Uses and Gratifications Theory. In *Proceedings of the 2nd International Conference on Control and Computer Vision* (Jeju, Republic of Korea) *(ICCCV 2019)*. Association for Computing Machinery, New York, NY, USA, 127–132. https://doi.org/10.1145/3341016.3341039

[61] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[62] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2003. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical care* 41, 11 (2003), 1284–1292.

[63] Ethan Kross, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. 2013. Facebook use predicts declines in subjective well-being in young adults. *PloS one* 8, 8 (2013), e69841.

[64] Reed Larson and Mihaly Csikszentmihalyi. 2014. *The Experience Sampling Method*. Springer Netherlands, Dordrecht, 21–34. https://doi.org/10.1007/978-94-017-9088-8_2

[65] Neal Lathia, Kiran K. Rachuri, Cecilia Mascolo, and Peter J. Rentfrow. 2013. Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 183–192. https://doi.org/10.1145/2493432.2493452

[66] H. Lee, Young Sang Choi, Sunjae Lee, and I. P. Park. 2012. Towards unobtrusive emotion recognition for affective social communication. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*. IEEE, New York, NY, USA, 260–264. https://doi.org/10.1109/CCNC.2012.6181098

[67] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (Taipei, Taiwan) *(MobiSys '13)*. Association for Computing Machinery, New York, NY, USA, 389–402. https://doi.org/10.1145/2462456.2464449

[68] Ruoyun Lin and Sonja Utz. 2015. The emotional responses of browsing Facebook: Happiness, envy, and the role of tie strength. *Computers in Human Behavior* 52 (2015), 29 – 38. https://doi.org/10.1016/j.chb.2015.04.064

[69] Ruoyun Lin, Niels van de Ven, and Sonja Utz. 2018. What triggers envy on Social Network Sites? A comparison between shared experiential and material purchases. *Computers in Human Behavior* 85 (2018), 271 – 281. https://doi.org/10.1016/j.chb.2018.03.049

[70] Jiangmeng Liu, Cong Li, Nick Carcioppolo, and Michael North. 2016. Do Our Facebook Friends Make Us Feel Worse? A Study of Social Comparison and Emotion. *Human Communication Research* 42, 4 (10 2016), 619–640. https://doi.org/10.1111/hcre.12090 arXiv:https://academic.oup.com/hcr/article-pdf/42/4/619/19494471/jhumcom0619.pdf

[71] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 368, 17 pages. https://doi.org/10.1145/3411764.3445467

[72] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 22 (March 2018), 26 pages. https://doi.org/10.1145/3191754

[73] Katerina Lup, Leora Trub, and Lisa Rosenthal. 2015. Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed. *Cyberpsychology, Behavior, and Social Networking* 18, 5 (2015), 247–252. https://doi.org/10.1089/cyber.2014.0560 arXiv:https://doi.org/10.1089/cyber.2014.0560 PMID: 25965859.

[74] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2014. Capturing the Mood: Facebook and Face-to-Face Encounters in the Workplace. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1082–1094. https://doi.org/10.1145/2531602.2531673

[75] Roberto Martinez-Pecino and Marta Garcia-Gavilán. 2019. Likes and problematic Instagram use: the moderating role of self-esteem. *Cyberpsychology, Behavior, and Social Networking* 22, 6 (2019), 412–416.

[76] Susan Y McGorry. 2000. Measurement in a cross-cultural environment: survey translation issues. *Qualitative Market Research: An International Journal* 3, 2 (2000), 74–81.

[77] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. Dancing to the Partisan Beat: A First Analysis of Political Communication on TikTok. In *12th ACM Conference on Web Science* (Southampton, United Kingdom) *(WebSci '20)*. Association for Computing Machinery, New York, NY, USA, 257–266. https://doi.org/10.1145/3394231.3397916

[78] Adrian Meier and Svenja Schäfer. 2018. The positive side of social comparison on social network sites: How envy can drive inspiration on Instagram. *Cyberpsychology, Behavior, and Social Networking* 21, 7 (2018), 411–417.

[79] Aske Mottelson and Kasper Hornbæk. 2016. An Affect Detection Technique Using Mobile Commodity Sensors in the Wild. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. ACM, New York, NY, USA, 781–792. https://doi.org/10.1145/2971648.2971654

[80] Yeamduan Narangajavana, Luis José Callarisa Fiol, Miguel Ángel Moliner Tena, Rosa María Rodríguez Artola, and Javier Sánchez García. 2017. The influence of social media in creating expectations. An empirical study for a tourist destination. *Annals of Tourism Research* 65 (2017), 60–70. https://doi.org/10.1016/j.annals.2017.05.002

[81] Evangelos Niforatos and Evangelos Karapanos. 2015. EmoSnaps: a mobile application for emotion recall from facial expressions. *Personal and Ubiquitous Computing* 19, 2 (2015), 425–444.

[82] Pawarat Nontasil and Stephen J. Payne. 2019. *Emotional Utility and Recall of the Facebook News Feed*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3290605.3300252

[83] Erik C Nook, Stephanie F Sasse, Hilary K Lambert, Katie A McLaughlin, and Leah H Somerville. 2018. The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological science* 29, 8 (2018), 1346–1357.

[84] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, et al. 2011. The impact of social media on children, adolescents, and families. *Pediatrics* 127, 4 (2011), 800–804.

[85] Charlotte Ottenstein and Tanja Lischetzke. 2020. Recall bias in emotional intensity ratings: investigating person-level and event-level predictors. *Motivation and Emotion* 44, 3 (2020), 464–473.

[86] Galen Panger. 2018. People Tend to Wind Down, Not Up, When They Browse Social Media. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 133 (Nov. 2018), 29 pages. https://doi.org/10.1145/3274402

[87] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[88] Matthew Pittman and Brandon Reich. 2016. Social media and loneliness: Why an Instagram picture may be worth more than a thousand Twitter words. *Computers in Human Behavior* 62 (2016), 155–167. https://doi.org/10.1016/j.chb.2016.03.084

[89] Marco Polignano, Fedelucio Narducci, Marco de Gemmis, and Giovanni Semeraro. 2021. Towards Emotion-aware Recommender Systems: an Affective Coherence Model based on Emotion-driven Behaviors. *Expert Systems with Applications* 170 (2021), 114382. https://doi.org/10.1016/j.eswa.2020.114382

[90] Daniel Preoţiuc-Pietro, H Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in facebook posts. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*. Association for Computational Linguistics, San Diego, California, 9–15.

[91] Katharina Rein and Tommaso Venturini. 2018. Ploughing digital landscapes: How Facebook influences the evolution of live video streaming. *New Media & Society* 20, 9 (2018), 3359–3380. https://doi.org/10.1177/1461444817748954 arXiv:https://doi.org/10.1177/1461444817748954

[92] Andreas Reissmann, Joachim Hauser, Ewelina Stollberg, Ivo Kaunzinger, and Klaus W. Lange. 2018. The role of loneliness in emerging adults' everyday use of facebook – An experience sampling approach. *Computers in Human Behavior* 88 (2018), 47–60. https://doi.org/10.1016/j.chb.2018.06.011

[93] Kat Roemmich and Nazanin Andalibi. 2021. Data Subjects' Conceptualizations of and Attitudes Toward Automatic Emotion Recognition-Enabled Wellbeing Interventions on Social Media. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 308 (oct 2021), 34 pages. https://doi.org/10.1145/3476049

[94] Morris Rosenberg. 2015. *Society and the adolescent self-image*. Princeton University Press, Princeton, N.J., USA.

[95] Philipp V. Rouast, Marc T. P. Adam, and Raymond Chiong. 2021. Deep Learning for Human Affect Recognition: Insights and New Developments. *IEEE Transactions on Affective Computing* 12, 2 (2021), 524–543. https://doi.org/10.1109/TAFFC.2018.2890471

[96] Mintra Ruensuk, Eunyong Cheon, Hwajung Hong, and Ian Oakley. 2020. How Do You Feel Online: Exploiting Smartphone Sensors to Detect Transitory Emotions during Social Media Use. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 150 (dec 2020), 32 pages. https://doi.org/10.1145/3432223

[97] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[98] Gillian M. Sandstrom, Neal Lathia, Cecilia Mascolo, and Peter J. Rentfrow. 2017. Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality* 69 (2017), 96–101. https://doi.org/10.1016/j.jrp.2016.06.004 Within-Person Variability in Personality.

[99] The New York Times SARAH PEREZ. 2009. New Illness: Facebook Depression? https://archive.nytimes.com/www.nytimes.com/external/readwriteweb/2009/02/02/02readwriteweb-new_illness_facebook_depression.html. Accessed: 2021-09-10.

[100] Zhanna Sarsenbayeva, Gabriele Marini, Niels van Berkel, Chu Luo, Weiwei Jiang, Kangning Yang, Greg Wadley, Tilman Dingler, Vassilis Kostakos, and Jorge Goncalves. 2020. Does Smartphone Use Drive Our Emotions or Vice Versa? A Causal Analysis. 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/10.1145/3313831.3376163

[101] Barbara Schneider. 2006. In the moment: The benefits of the experience sampling method. In *The work and family handbook*. Lawrence Erlbaum Associates, NJ, USA, 469–488.

[102] Edmund Seto, Jenna Hua, Lemuel Wu, Aaron Bestick, Victor Shia, Sue Eom, Jay Han, May Wang, and Yan Li. 2014. The Kunming CalFit study: Modeling dietary behavioral patterns using smartphone data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, New York, NY, USA, 6884–6887. https://doi.org/10.1109/EMBC.2014.6945210

[103] Nicolas Simonazzi, Jean-Marc Salotti, Caroline Dubois, and Dominique Seminel. 2021. Emotion Detection Based on Smartphone Using User Memory Tasks and Videos. In *Human Interaction, Emerging Technologies and Future Applications III*, Tareq Ahram, Redha Taiar, Karine Langlois, and Arnaud Choplin (Eds.). Springer International Publishing, Cham, 244–249.

[104] Andrew D. Smock, Nicole B. Ellison, Cliff Lampe, and Donghee Yvette Wohn. 2011. Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior* 27, 6 (2011), 2322–2329. https://doi.org/10.1016/j.chb.2011.07.011

[105] Peta Stapleton, Gabriella Luiz, and Hannah Chatwin. 2017. Generation validation: The role of social comparison in use of Instagram among emerging adults. *Cyberpsychology, Behavior, and Social Networking* 20, 3 (2017), 142–149.

[106] Mai-Ly N Steers, Robert E Wickham, and Linda K Acitelli. 2014. Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms. *Journal of Social and Clinical Psychology* 33, 8 (2014), 701–731.

[107] Anika Stuppy, Nicole L Mead, and Stijn M J Van Osselaer. 2019. I Am, Therefore I Buy: Low Self-Esteem and the Pursuit of Self-Verifying Consumption. *Journal of Consumer Research* 46, 5 (06 2019), 956–973. https://doi.org/10.1093/jcr/ucz029 arXiv:https://academic.oup.com/jcr/article-pdf/46/5/956/31700831/ucz029.pdf

[108] Edson C. Tandoc, Patrick Ferrucci, and Margaret Duffy. 2015. Facebook use, envy, and depression among college students: Is facebooking depressing? *Computers in Human Behavior* 43 (2015), 139 – 146. https://doi.org/10.1016/j.chb.2014.10.053

[109] Marika Tiggemann and Belinda McGill. 2004. The role of social comparison in the effect of magazine advertisements on women's mood and body dissatisfaction. *Journal of Social and Clinical Psychology* 23, 1 (2004), 23–44.

[110] Ricvalerio Valério. 2017. ricvalerio/foregroundappchecker. https://github.com/ricvalerio/foregroundappchecker

[111] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *ACM Comput. Surv.* 50, 6, Article 93 (Dec. 2017), 40 pages. https://doi.org/10.1145/3123988

[112] M Van den Brink, ENG Bandell-Hoekstra, and H Huijer Abu-Saad. 2001. The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. *Headache: The Journal of Head and Face Pain* 41, 1 (2001), 11–20.

[113] Gert Van Dijck and Marc M. Van Hulle. 2006. Speeding Up the Wrapper Feature Subset Selection in Regression by Mutual Information Relevance and Redundancy Analysis. In *Artificial Neural Networks – ICANN 2006*, Stefanos D. Kollias, Andreas Stafylopatis, Włodzisław Duch, and Erkki Oja (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 31–40.

[114] Philippe Verduyn, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General* 144, 2 (2015), 480.

[115] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, and Markus Gross. 2020. Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data. 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, 1–13. https://doi.org/10.1145/3313831.3376504

[116] Georgia Wells, Jeff Horwitz, and Deepa Seetharaman. 2021. *Facebook knows Instagram is toxic for teen girls, company documents show.* The Wall Street Journal. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739

[117] K. Yang, C. Wang, Y. Gu, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, and J. Goncalves. 2021. Behavioral and Physiological Signals-Based Deep Multimodal Approach for Mobile Emotion Recognition. *IEEE Transactions on Affective Computing* Preprint, 01 (Jul 2021), 17 pages. https://doi.org/10.1109/TAFFC.2021.3100868

[118] Xiao Zhang, Wenzhong Li, Xu Chen, and Sanglu Lu. 2018. MoodExplorer: Towards Compound Emotion Detection via Smartphone Sensing. *Proc. ACM*

Interact. Mob. Wearable Ubiquitous Technol. 1, 4, Article 176 (Jan. 2018), 30 pages. https://doi.org/10.1145/3161414

[119] Xiao Zhang, Fuzhen Zhuang, Wenzhong Li, Haochao Ying, Hui Xiong, and Sanglu Lu. 2019. Inferring Mood Instability via Smartphone Sensing: A Multi-View Learning Approach. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 1401–1409. https://doi.org/10.1145/3343031.3350957

[120] S. Zhao, H. Yao, Y. Gao, G. Ding, and T. Chua. 2018. Predicting Personalized Image Emotion Perceptions in Social Networks. *IEEE Transactions on Affective Computing* 9, 4 (Oct 2018), 526–540. https://doi.org/10.1109/TAFFC.2016.2628787

[121] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (2017), 30–50. https://doi.org/10.1016/j.sheji.2017.08.003