# What Life Events are Disclosed on Social Media, How, When, and By Whom?


Koustuv Saha
Georgia Institute of Technology
Atlanta, GA, USA
koustuv.saha@gatech.edu

Jordyn Seybolt
Georgia Institute of Technology
Atlanta, GA, USA
jordynseybolt@gatech.edu

Stephen M. Mattingly
University of Notre Dame
South Bend, IN, USA
smattin1@nd.edu

Talayeh Aledavood
Georgia Institute of Technology
Atlanta, GA, USA
talayeh@gatech.edu

Chaitanya Konjeti
Georgia Institute of Technology
Atlanta, GA, USA
ckonjeti1@gatech.edu

Gonzalo J. Martinez
University of Notre Dame
South Bend, IN, USA
gmarti11@nd.edu

Ted Grover
University of California, Irvine
Irvine, CA, USA
grovere@uci.edu

Gloria Mark
University of California, Irvine
Irvine, CA, USA
gmark@uci.edu

Munmun De Choudhury
Georgia Institute of Technology
Atlanta, GA, USA
munmund@gatech.edu



## ABSTRACT

Social media platforms continue to evolve as archival platforms, where important milestones in an individual's life are socially disclosed for support, solidarity, maintaining and gaining social capital, or to meet therapeutic needs. However, a limited understanding of how and what life events are disclosed (or not) prevents designing platforms to be sensitive to life events. We ask what life events individuals disclose on a 256 participants' year-long Facebook dataset of 14K posts against their self-reported life events. We contribute a codebook to identify life event disclosures and build regression models on factors explaining life events' disclosures. Positive and anticipated events are more likely, whereas significant, recent, and intimate events are less likely to be disclosed on social media. While all life events may not be disclosed, online disclosures can reflect complementary information to self-reports. Our work bears practical and platform design implications in providing support and sensitivity to life events.



## CCS CONCEPTS

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

## KEYWORDS

social media, life events, language, self-disclosure, audience, self-reports, individual differences




*CHI '21, May 8–13, 2021, Yokohama, Japan*



ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445405

**ACM Reference Format:**
Koustuv Saha, Jordyn Seybolt, Stephen M. Mattingly, Talayeh Aledavood, Chaitanya Konjeti, Gonzalo J. Martinez, Ted Grover, Gloria Mark, and Munmun De Choudhury. 2021. What Life Events are Disclosed on Social Media, How, When, and By Whom?. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3411764.3445405


## 1 INTRODUCTION

> "Yes, my year looked like that. True enough. My year looked like the now-absent face of my little girl. It was still unkind to remind me so forcefully. [..] The Year in Review ad [kept] coming up in my feed, rotating through different fun-and-fabulous backgrounds, as if celebrating a death, and there [was] no obvious way to stop it" – Eric Meyer [102]

Ups and downs are inevitable in an individual's life. As social media platforms continually emerge as important parts of many of our lives [121], they serve many needs and purposes surrounding those very ups and downs of life. Not only do these platforms enable individuals to connect with others and share day-to-day happenings in life [17, 59, 156], they also have explicit affordances [21] in design that allow individuals to record and archive their important life events. For instance, the Facebook timeline reminds people of birthdays and personal milestones.

Toward better user experience, most social media platforms today employ algorithms to recommend, rank, or curate personalized content. However, despite providing affordances to gather information on life events, social media content personalization largely relies on topics, interests, and social connections, and rarely accounts for an individual's life events. For this reason, when a Facebook user Eric Meyer was shown his "Year in Review" on the platform in 2014 that included his now-dead daughter's picture, he felt the feature to not only be jarring but also emotionally triggering – labeling the News Feed algorithm as "inadvertently cruel" due to its insensitivity to people's life events [102].

We note that in their attempts to serve as safe spaces for authentic expression, support seeking, and promoting wellbeing [23, 37], social media platforms need to consider affordances and algorithms that are sensitive to, respectful of, and compassionate towards major happenings in an individual's life. Such an approach can improve the value one can gain from social media participation, such as meeting varied emotional, informational, and therapeutic needs, and empowering people to gain, maintain, and leverage their social capital. Furthermore, research in human-computer interaction (HCI) and computer-mediated communication (CMC) reveals how naturalistic, self-initiated, and open-ended forms of social data recording, enabled by social media, can augment our understanding of people's reactions and behavior changes surrounding major life events, such as gender transition [66], death of a loved one [19, 97], child birth [36], job loss [23], and pregnancy loss [7]. For example, after a personal crisis, people may desire to reach out to their social media networks for support [7], and following a job loss, an individual may seek empathy from their online social ties and seek new opportunities or job search-related resources from their weak ties [23]. **Together, this calls for a critical need to understand social media disclosures of life events.**

A life event disclosure on social media uniquely conveys how someone perceives and shares their feelings about the event. However, from an individual's perspective, deciding to self-disclose something as sensitive as a life event on social media can be influenced and compounded by various factors. Literature outlines social media disclosure is affected by factors related to self-presentation, social desirability, audience, boundary regulation, and stigma — people may want to be viewed in particular ways across different audiences, or may not be comfortable about sharing some aspects of their lives with their social media audience [54, 80, 96]. Importantly, an individual may not disclose all life events on social media, and the disclosure choices may vary across individuals and situations. However, the specific factors that explain disclosures (and non-disclosures) remain largely unknown. A deeper examination of life event disclosures would help us understand the authenticity of social media postings regarding how closely this data reflects real-world occurrences of life events in one's life. This would also help to design platform affordances that account for and are sensitive to an individual's life events, and content curation/recommendation algorithms that more adequately represent the gap between observed and unobserved social media behaviors.

Towards designing platforms sensitive to life events, this formative study seeks to understand what life events people disclose or withhold on social media, how and when these disclosures happen, and what are the attributes of individuals who tend to disclose versus not. To accomplish our research goal, in the absence of "true ground-truth" of life event occurrences, we compare social media disclosures of life events with life events self-reported on a standardized survey. Specifically, we use year-long Facebook data from 236 participants who also responded to a retrospective survey, adapted from the PERI life events scale [42], which inquired about life event occurrences in the past year. We ask the below research questions:

**RQ 1:** How do social media self-disclosures of life events deviate from self-reported survey?

**RQ 2:** How do individual and event attributes explain the deviation in life event disclosure on social media compared to the self-reported survey?

First, targeting the question of *how online self-disclosures of life events deviate from self-reports*, we qualitatively code and define life event disclosures on Facebook data. **Our work contributes a comprehensive codebook (available for theory and practice) that enhances our understanding of social media disclosures of life events.** We thematically analyze the language of life event descriptions on social media as compared to their occurrences, with insightful findings such as: social media life event disclosures are typically expressive and emotional in nature; multiple life events may be recorded in the two modalities – social media and survey that might be related, unrelated, or causal; and that negative events tend to stand out in the retrospective recall of individuals, manifested through their survey responses.

Second, given an individual and a life event, we examine *how individual attributes (demographics and traits) and event attributes explain the deviation in disclosure on social media compared to self-reports*. We build logistic regression models of logging behaviors by controlling for individual and event attributes. Here individual attributes correspond to demographics and intrinsic traits of cognitive ability, personality, and affect, and event-centric attributes correspond to valence, significance, recency, anticipation, intimacy, scope, status and type of event. Our analyses reveal significant findings advancing our understanding of online life event disclosures: positive and anticipated events are more likely to be disclosed online, whereas significant, recent, and intimate events bear a propensity to be self-reported in survey.

Our findings reveal how different life events may elicit varied decision-making processes on the part of social media users surrounding what, when, and how to disclose, while also navigating the underlying norms of the platform and the audience of a potential disclosure. Then by unpacking the fundamental differences between social media platforms and surveys as it pertains to their respective context of use and available affordances, we discuss a need to understand and straddle the socio-technical gap [1] between what individuals disclose online in a self-initiated, intrinsically motivated manner, and what they self-report offline to a prompted survey conducted by a more private but unfamiliar audience of researchers. Drawing on these theoretical underpinnings and implications, we argue that a "one size fits all" approach to scaffold online life event disclosures may not work. We conclude by providing design suggestions for social computing systems that are sensitive to people's life events, including strategies that accommodate non-disclosure practices and that provide agency to those social media users who choose not to disclose specific life events.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Life Events: Importance and Assessment

Harkness and Monroe define life events as "environmental changes that have a definable beginning point in time and that would be expected to be associated with at least some degree of psychological threat, unpleasantness, or behavioral demands." Life events have varying importance, severity, and valence [42]. Acute and major life events require substantial behavioral adjustments and can cause physical and psychological distress [152]. They are, therefore, a predictor of various physical and mental conditions such as chronic fatigue [137], depression [2, 149], and anxiety [83]. The effect of these events on people's lives has long been a topic of study by social scientists. This has led to designing and implementing methodologies to identify and assess major life events [41, 73, 140].

Life events are predominantly assessed with survey questionnaires and interviews [103]. These assessments are typically conducted after a period of time, such as during longitudinal assessment of wellbeing, individuals are inquired about life events they encountered in the last N days [41, 56, 73]. Alternatively, life events are also inquired as a part of other psychological assessments following a major crisis or a stressful event [32]. These approaches typically include a checklist of major life events, and individuals respond to the items that best relate to them and describe them with significance and valence of the effect on them [41]. While these methods are efficacious, reliable, and validated, these assessments are largely conducted after the respondents are displaced in space and time from the event occurrences [108]. This may lead to retrospective recall and hindsight bias; individuals are more likely to report experiences that seem personally more relevant, occurred more recently, stand out as significant or unusual, or those more consistent with current mood states [154]. Further, recollecting stressful events from the past can cause a respondent to undergo similar trauma associated with the event, and conducting these delicate surveys can be hard in sensitive circumstances [139].

Consequently, research has encouraged in-the-present forms of data recording, such as experience sampling and journaling [157, 158]. These approaches can not only capture short-term, yet valuable dynamics, but can also elicit positive effects on the individual for being expressive about life experiences [69]. In recent times, social media is considered to be a similar form of in-the-present data where individuals feel an intrinsic motivation to record, archive, and share life experiences in naturalistic settings [23, 66]. However, recording life events on social media is compounded by factors such as social desirability, self-presentation, and privacy [96]. It remains relatively unknown, who would be comfortable to record what kinds of life events on social media — the key question explored in this work. We examine how individual-centric and event-centric attributes explain social media disclosures of life events.

## 2.2 Self-disclosure and Public Audience

Jourard defined self-disclosure as "the act of revealing personal information to others" [79]. Self-disclosures about experiences and thoughts comprise a substantial part (approximately 30-40%) of what people share with others [45, 90]. Tamir and Mitchell showed that self-disclosures tend to tie to intrinsic values for individuals and therefore, are rewarding [147]. However, sharing about oneself comes with risks such as vulnerability, lower control, and losing privacy [4, 13, 120]. Omarzu noted that breadth, duration, and depth of self-disclosure are functions of subjective utility and risks for a self-disclosing individual [112]. At a neurophysiological level, a recent study shows that the degree of self-disclosureassociates with intrinsic functional connectivity of certain brain regions [101]. According to Goffman's self-presentation theory, people desire to control the impressions they give to others and therefore manage the impressions through social performance [54, 87]. Goffman used the notions of "frontstage" and "backstage" — frontstage refers to the appearance we put on for the public and backstage refers to the personal space where people do the necessary work to give the desired impressions on the front stage [54].

Compared to face-to-face interactions, online self-disclosures tend to make up for an even larger fraction of what people share with others in computer mediated communication (CMC) such as that on social media [77, 106]. Here, people can have more control on presenting themselves [46, 87]. These platforms can act as a space where people manage impressions and showcase their "best self" and therefore use it as a front stage in Goffman's terminology [71, 100, 114, 155]. Social media can also act as a back stage because they are access controlled [71]. It is a space where people get to see certain sides of others that they would not get to see in the physical world [71, 105, 128]. Building on Goffman's theory, Hogan argued that social media disclosures have properties of an "exhibition" rather than a "performance" — what people share on these media is seen as artifacts which are archived in databases (storehouses) and are shown to the friends and followers (audience) based on the algorithms and the means that the platform provides for presenting the data (curator) [71].

Major motivations of "public self-disclosure" on social media correspond to the opportunities to self-broadcast and to build personal connections with others [13]. Kim et al. examined the motivations of posting selfies on social media by adopting the theory of planned behavior [3], finding that attitude towards selfie-posting, subjective norms, perceived behavioral control, and narcissism are key factors contributing to the act of selfie posting [87]. Prior work also studied who discloses what on social media based on individual differences [72, 88, 141]. Sheldon compared self-disclosure for males and females with different types of friendships, finding that self-disclosure to recently added Friends is higher for males, whereas self-disclosure to exclusively Facebook friends and exclusively face-to-face friends is higher for females [141]. Another study found that, females express more positive emotions on Twitter than males [88]. Other individual differences such as age and personality traits have also been noted to explain self-disclosure on social media [72, 150].

Similar to the physical world, online self-presentation is influenced by the audience [14, 96]. The public facing nature of social media platforms can increase an individual's accountability and reduce deception in online spaces [44, 62, 124]. Prior work revealed that social media facilitates candid self-disclosure [37, 125], and unique affordances such as anonymity, throwaway accounts, and selective audiences enhance self-disclosure of life events and experiences [8, 9, 159]. Research also noted the positive benefits of online self-disclosure [8, 37, 78, 148], such as in decreasing loneliness [148] and increasing life satisfaction [150]. In online communities, individuals feel a sense of belonging, and seek solidarity during stressful circumstances [95, 130]. Relatedly, Ernala et al. adopted the Social Penetration Theory to operationalize intimacy in self-disclosure and studied the therapeutic benefits of stigmatized self-disclosure on Twitter [48]. Together, prior work motivates us in studying self-disclosure of life events on social media, particularly to compare and contrast disclosure to an online (semi-) public audience versus self-reports in an offline private audience.

## 2.3 Disclosure of Life Events on Social Media

As part of self-disclosure on social media, people share their life events on these platforms. Prior work has looked at major life events, transitions, and important markers for individuals [7, 36, 38, 64, 66, 131]. De Choudhury et al. examined social media behavior changes around a major life event, particularly postpartum changes in behavior and mood of new mothers along the dimensions of social engagement, emotion, social network, and linguistic style [35]. Social media has also enabled disclosures of sensitive and stigmatized life events, such as gender transitions [66, 67] and pregnancy loss [7]. Burke and Kraut studied how individuals interact with strong and weak ties of friendships on Facebook following a loss of job [23]. Andalibi and Forte proposed a decision framework to understand six types of decision factors related to disclosing pregnancy loss on social media: self-related, audience-related, societal, platform and affordance-related, network-level, and temporal [7]. Andalibi followed this up with the complementary question of examining the factors that lead to *non-disclosures* of pregnancy-loss on Facebook [6].

Relatedly, Massimi and Baecker studied how family members use technologies to remember their loved ones [97]. Prior research has also studied how individuals disclose about death of close ones [20, 52]. Other longitudinal studies have examined behavioral changes with respect to exogenous or endogenous, anticipated or unanticipated events, e.g., antidepressant use [133], alcohol and substance use [86, 93], and diagnosis with health conditions [49, 63]. Bevan et al. studied the difference in sharing different types of positive and negative life events in directness on Facebook [15]. Our work extends this body of work by providing a deeper understanding of what life events people choose to disclose (or not disclose), adopting a comprehensive list of various life events.

In parallel, researchers have conducted computational studies to extract and analyze life event disclosures on social media [40, 91, 167], and a recent work analyzed leakage of privacy in life event disclosures on Twitter [82]. A major challenge in these studies has been that life event disclosures only constitute a small fraction of all kinds of social media posts. Together, we note that a majority of studies have either adopted broad definitions of life events for automatic identification or have focused on very specific life events [25, 28, 85]. Our work aims to bridge this gap by adopting a theoretical lens to investigate life event disclosures on social media, and by contributing a comprehensive and fine-grained codebook to identify life events on social media.

# 3 STUDY AND DATA

This paper uses data from a large-scale longitudinal study called Tesserae [99]. This study recruited 754 participants who are information workers in cognitively demanding fields in diverse job positions and roles (e.g., engineers, consultants, and managers) at various organizations in spread across the U.S. The project broadly aims to study wellbeing by leveraging multiple modalities of data. The participants were enrolled between January 2018 and July 2018, and were requested to remain in the study for either up to a year or through April 2019. The participants either received a series of staggered stipends totaling $750 or they participated in a set of weekly lottery drawings (multiples of $250 drawings) depending on their employer restrictions.

**Privacy and Ethics.** The Tesserae project was approved by the Institutional Review Board at the researchers' institutions. Given the sensitivity of the data, participant privacy was a key concern. The participants were provided with informed-consent documents describing the specifics of what data they were providing, and how would that be stored. The participants needed to consent to each form of data, and could also clarify concerns and opt out of any data collection. The data was de-identified and stored in secured databases and servers which were physically located in the researcher institutions, and had limited access privileges.

## 3.1 Social Media Data

The Tesserae project asked consented participants to authorize their social media data, particularly Facebook, *unless they opted out, or did not have an account*. The enrollment briefing and consent process explicitly explained that their study participation did not necessitate them to use social media in a particular fashion, and they were expected to continue with their typical social media use. Participants authorized access to their social media data through an Open Authentication (OAuth) based data collection infrastructure developed in Saha et al. [129]. OAuth protocol is an open standard for access delegation, commonly used as a way for internet users to log in and grant third party access to their information, without sharing passwords. The OAuth protocol provides a more privacy-preserving and convenient means of data collection at scale, over secured channels without the transfer of any personal credentials.

Given that Facebook is the most popular social media platform [58] and its longitudinal nature has enabled social media studies of individual differences [7, 36], it suits our problem setting of understanding life event disclosures on social media. Facebook is also the most prevalent social media stream in the Tesserae dataset. Among these, the total 572 participants who provided access to Facebook data, 242 participants did not make any update during the year-long study period between January 2018 and April 2019 — the same period when the participants' self-reported life event occurrences were also collected. This paper uses a 14,202 posts data of the remaining 330 participants to identify life event disclosures in Section 4.1, which was followed by examining factors for life event disclosures on a subset of 236 participants' data who also responded to self-reported survey on life events, explained below.

## 3.2 Self-Reported Survey Data

*3.2.1 Start of Participation Period: Data on Individual Differences and Psychological Traits.* The enrollment process consisted of an initial survey questionnaire related to demographics (age, gender, education, etc.), and survey questionnaires of self-reported psychological constructs, including: 1) *Cognitive Ability*, as assessed by the Shipley scales of Abstraction (fluid intelligence) and vocabulary (crystallized intelligence) [142], 2) *Personality Traits*, the big-five personality traits as assessed by the Big Five Inventory (BFI-2) scale [143, 151], and 3) *Wellbeing*, the general positive and negative affect levels as assessed through the Positive And Negative Affect (PANAS-X) scale [164], the anxiety level as measured via State Trait scale [144], and the quality of sleep as measured via the Pittsburg Sleep Quality Index (PSQI) scale [43]. These scales allow to capture individual differences that may modulate a user's choice and preferences about reporting a life event on social media and on survey. Table 1 summarizes the distribution of the self-reported data within our dataset, where we find a reasonably well distribution in demographics and psychological traits among our participants.

*3.2.2 End of Participation Period: Life Events Survey Data.* At the end of the participation period of the Tesserae study, participants were optionally asked to fill in a life events survey. We designed this life events survey drawing on the Psychiatric Epidemiology Research Interview (PERI) life events scale [41]. Life events were broadly categorized as School, Personal, Work/Organization, Health, Financial, Local/Regional, and Other. For each category, the survey also included example seed events to help the participant understand respective categories. Participants were briefed that they could refer to their calendars and any relevant personal diaries or journals while completing the survey, to verify the events and dates. The survey was designed in such a way that participants could enter more than one event, and include corresponding attributes about the events. These attributes include a brief description of the event, and two 7-item Likert scales of self-identified significance (Lowest to Highest significance) and valence (Extremely Negative to Extremely Positive) of the life event. In addition, participants entered the start and end date range, status of the event (ongoing or ended), and a confidence value (7-item Likert scale from Lowest to Highest Confidence) regarding the occurrence of the event. Table 2 shows the different categories of life events in our survey along with category hint provided in survey and example self-reported descriptions from the responses.

Out of the initial total of 754 participants, 423 participants responded to these surveys with 1,547 entries of life events during the study participation period (mean = 3.86 events per individual). Out of these 423 responded participants, 236 provided us the social media data (above subsection). We examine the data of these 236 participants to understand the deviation of online self-disclosure of life events from self-reports.

# 4 METHODS

## 4.1 Defining and Annotating Life Event Disclosures on Social Media

Social media facilitates self-disclosures of feelings and experiences from day-to-day lives [8, 48]. From a standpoint of life event disclosures, social media posts are unstructured forms of textual expressions, and this data lacks "ground-truth" labels regarding what constitutes a life event disclosure and what does not. So we first aim to systematically identify online self-disclosures of life events from social media data with respect to a theoretical grounding of life event occurrences. We adopt a qualitative coding approach to iteratively define and annotate life event expressions on social media. We primarily build on and adapt the list of categories from the PERI life event scale [41] in the context of social media data. Our theory-driven coding enables us to formally define a social media post to contain a life event disclosure *if the post describes an event which is directly or indirectly associated with the individual or their close ones, such that it potentially leaves a psychological, physiological, or behavioral impact, or be significant enough to be remembered after a period.* This section first explains our annotation approach, followed by our examinations to study the deviation of life event disclosures on social media as compared to self-reported surveys.

While the PERI life events scale [42] identified a list of various life event categories, there is no established means to adopt this on social media data. Therefore, we applied these categories in a kind of directed coding approach [74], i.e., when developing the codebook we also allowed concepts and meanings to emerge from posts in somewhat of an open coding [146]. Our codebook is particularly driven towards identifying life event disclosures from social media language. The Supplementary File provides the detailed codebook to identify life event disclosures on social media.

We recruited five annotators who are undergraduate students. Although our Facebook data primarily consists of English posts and belongs to a participant pool recruited in the U.S., all participants were demographically and culturally heterogeneous. Therefore, it is important to note that our annotators (three women and two men) belonged to diverse cultural backgrounds; in race/ethnicity, two identified as Caucasian, two as East Asian, and one as South Asian. During discussions, we found specific occurrences when annotators were able to identify culturally significant events due to their cultural backgrounds, which could have been missed by other annotators. These five annotators first coded a random sample of 140 Facebook posts with the PERI life events scale [42] and the instruction that they could add new categories if a post was a life event disclosure and it did not fit any of the existing PERI categories.

The annotators and two authors then discussed the coding one by one in detail. Together, we made decisions on all posts with coding discrepancies, and revised our codebook based on agreeable themes. These included resolving boundary and similar sounding cases such as identifying a *trip* versus a *vacation*. Next, the annotators separately coded an additional 50 randomly selected posts. For the total 200 posts, we found a high agreement of 88% between the annotators and an average Fleiss $\kappa$ of 0.71. Two annotators then independently coded the remaining 14,002 posts. Because of the subjectivity in social media data, we adopted a liberal identification strategy that a post is labeled as a self-disclosure of life event if it is labeled so by either of the two annotators. We discussed several explicit and boundary cases to decide general criteria for identifying life event disclosures, which we elaborate on in Table 3. We note that the presence of a post within the context of other posts (before and after it) drove our decision-making towards labeling a post.

## 4.2 Comparing Life Events Disclosed on Social Media Versus Reported on Survey

So far, we described our approach to obtain life events disclosed on social media and self-reported in surveys. Consequently, for the common set of 236 participants for whom we have both modalities of data, we obtain 912 life events self-reported on the survey and 1,669 self-disclosed on social media. To answer our core research question on *what, how, when, and by whom* life events are disclosed on social media compared to self-reported surveys, first, we examine the distribution of life events in the two modalities of datasets. Then, we conduct a thematic analysis of the overlapping life event

**Table 1: Descriptive statistics of self-reported demographics and psychological constructs of 236 participants with both social media and life events survey data.**

| Covariates | Value Type | Values / Distribution |
|---|---|---|
| *Demographic Characteristics* | | |
| Gender | Categorical | Male ($n$=121) \| Female ($n$=115) |
| Born in U.S. | Categorical | Yes ($n$=218) \| No ($n$=18) |
| Age | Continuous | Range (22:63), Mean = 36.57, Std. = 9.88 |
| Education Level | Ordinal | 5 values [HS., College, Grad. Student, Master's, Doctoral] |
| *Cognitive Ability (Shipley scale)* | | |
| Fluid (Abstraction) | Continuous | Range (5:24), Mean = 16.93, Std. = 2.94 |
| Crystallized (Vocabulary) | Continuous | Range (22:40), Mean = 33.70, Std. = 3.32 |
| *Personality Trait (BFI scale)* | | |
| Openness | Continuous | Range (2.17:5), Mean = 3.84, Std. = 0.61 |
| Conscientiousness | Continuous | Range (1.92:5), Mean = 3.94, Std. = 0.63 |
| Extraversion | Continuous | Range (1.67:4.92), Mean = 3.42, Std. = 0.68 |
| Agreeableness | Continuous | Range (2.25:5), Mean = 3.95, Std. = 0.55 |
| Neuroticism | Continuous | Range (1:4.58), Mean = 2.44, Std. = 0.78 |
| *Affect and Wellbeing* | | |
| Pos. Affect | Continuous | Range (13:49), Mean = 34.24, Std. = 5.69 |
| Neg. Affect | Continuous | Range (10:37), Mean = 16.83, Std. = 4.62 |
| Anxiety | Continuous | Range (20:67), Mean = 37.83, Std. = 9.33 |
| Sleep Quality | Continuous | Range (1:16), Mean = 6.80, Std. = 2.57 |

logs from the two datasets. Finally, we examine the factors that explain the overlap and deviation in reportage on either or both the modalities, for which, we describe the statistical tests in the following subsection.

## 4.3 Examining Factors Associated with Life Events Disclosures and Survey Self-Reports

To examine the factors explaining deviation in recording life events on the two modalities, we first identify a set of theory-driven covariates that may contribute to an individual's life event disclosure (or no disclosure) on either or both the modalities. We then use these covariates in our statistical tests and models to explain such life event disclosure.

*4.3.1 Covariates.* Given an individual and a life event, our covariates belong to two major kinds — *individual centric attributes* and *event-centric attributes*, which we describe below.

*Individual-centric Attributes.* Given that an individual's disclosure is known to be driven by their demographic and intrinsic traits, we use individuals' demographic and psychological attributes (as in Table 1) in our models.

**Demographics.** Prior studies controlled on several demographic attributes in studying self-presentation and self-disclosure of individuals [132]. We include demographic variables of gender, age, born in the U.S., educational level, and income in our models.

**Cognitive Ability.** Cognitive ability is known to associate with an individual's disclosure and expressiveness [123], which we include as independent variables in our model. We used the the Shipley scales of 1) Abstraction measuring fluid cognitive ability and 2) Vocabulary measuring crystallized cognitive ability (Section 3) [142].

**Personality.** Prior work revealed the role of personality in people's disclosure, including in online settings [72, 138]. We include personality trait as a covariate in our models where ground-truth assessments of personality traits come from the Big-Five inventory along the traits of openness, conscientiousness, extraversion, agreeableness, and neuroticism [143].

**Affect and Wellbeing.** Social media use is known to be associated with people's trait based measures of affect and wellbeing [162]. We include positive and negative affect traits as assessed by the PANAS-X scale [164], anxiety trait as assessed by the STAI-Tait scale [144], and sleep quality as assessed by the PSQI scale [43]. We note that PSQI scale assesses sleep quality in such a way that lower values indicate healthier sleep. Therefore, for easier interpretation, we reverse-scale the values and use "Healthy Sleep Quality" as a covariate which directly correlates with healthier sleep.

*Event-centric Attributes.* People's life event disclosures (or non-disclosures) may be driven by event-centric attributes. We describe the motivation and the operationalization of event-centric attributes considered in our models below.

**Event Recency.** Self-reported surveys are known to be biased to more recent events [10, 53]. However, no such evidence exists about social media postings, which is more of a self-initiated and in-the-present recording. To understand such an effect in online life event disclosure, we include recency of events as an independent variable. We first choose a reference date as the date of conducting the end of participation survey. Then, for the survey data, we calculate the number of days between the reference date and the self-reported occurrence of event (also collected in the survey: Section 3). For social media data, we calculate the number of days between the reference date and the date of posting. For easier interpretation and standardization, we reverse-scale the number of dates to obtain recency on a min-max scale of 0 to 1 — such that 1 represents most recent event whereas 0 represents least recent events.

**Event Significance.** Individuals are known to be more likely to recall and report events which bear greater degree of significance in their lives in whatsoever ways [107]. This aligns with survival

**Table 2: Life Event categories, example hints provided in survey, and example self-reported description in the post-participation self-reported survey — survey scale drawn on the PERI life events scale [41].**

| Event Type | Category Hint | Example Self-Reported Description |
|---|---|---|
| School | Back to school, Changed school, Finished school, Issue at school, etc. | Accepted to business school |
| Personal | Getting married or divorced, Having a child, Experiencing a death of someone close, Moved residences, Damage of property, etc. | Was working on an adoption |
| Work | Changed jobs, Received a promotion, Was fired, Had performance review, Received bonus, End of quarter or year, Reorganization | Given more responsibilities in my job, which made me realize I don't want to work in this job anymore |
| Health | Physical illness or injury, Health treatment, Miscarriage/Stillbirth, Pregnancy related changes, Started menopause, Health changes | Mother diagnosed with kidney failure and congestive heart failure |
| Financial | Went into debt, Took out mortgage, Made a large purchase, e.g. car or home, Experienced financial gain or loss | Paid off 2 vehicles and refinanced one to pay off high interest credit cards |
| Local/ Regional | Weather-related changes (blizzard, flood, storm, etc.), Societal changes (political or economic event, sports event, mass-shooting, etc.) | Was at a baseball game where my team advanced to National League Championship |
| Other | Any other events that do not fall under the above categories | - |

salience [107], and emotional or informational relevance can drive the salience in memory [84, 119]. Participants self-reported how significant they considered each life event they logged — which we use as an independent variable for event records from surveys. For events recorded on social media, we adopt the significance rating per event as per the PERI life events scale [41]. We separately standardize the significance scores on a min-max scale of 0-1 to make the significance scores comparable across the modalities, and then use this scaled score as an independent variable in our models.

**Valence.** Our independent variables include valence or sentiment of the event, in terms of being positive or negative. Like above, valence directly associates with emotional relevance of an event in the memory [84]. The survey data included people's self-reported valence on a Likert scale of extremely positive to extremely negative, which we group into three bins of positive, neutral, and negative to minimize subjectivity in our analyses. To score valence of social media life events, we use the VADER tool [75] to identify the major sentiment of a post among positive, negative, and neutral, which we use as the valence for life event entries from social media data.

**Anticipation of an Event.** Life events include a characteristic on the basis of anticipation: Compas et al. defined anticipated events as the events which an individual can either hope or worry about in the next six months [32]. We adopt a similar definition to label each life event in our dataset with binary labels of anticipated or unanticipated. Example anticipated events are buying a house, childbirth/pregnancy related events, whereas example unanticipated events are accidents or getting fired from work.

**Intimacy in Disclosure.** Prior work studied that intimacy is a core attribute that might moderate people's disclosure behavior [5, 49, 94]. Intimacy relates to the degree to which one can comfortably open up about a particular event at personal, close, trusted others, and public circles of relationships [49, 54]. While social media disclosures are broadcasted to some form of public or known private audience, a self-reported survey is likely self-perceived to be much more private. We draw upon the annotation scheme from Ernala et al. to code life event descriptions — we annotate both survey self-reports and online disclosures of life events on a degree of intimacy Likert scale of Low, Medium, and High[1].

**Scope of an Event.** The social ecological model posits that an individual's wellbeing is impacted by different layers of scope ranging across individual, relationships with close ones, societal, and local factors [24]. Similarly, the scope of a life event can either be directly on the individual themselves, or their close ones, or something more generic [15]. We label each life event in our datasets with their ecological scope of directness on a three-point Likert scale[1] such that 1) *Low* scope events include generic events such as bad weather or neighborhood related events, 2) *Medium* scope events are associated with someone close and leave an indirect effect on the individual (e.g., spouse's pregnancy, child going to school), and 3) *High* scope events are unique to and direct on the individual, e.g., being fired from job themselves.

**Temporal Status.** We also include temporal status of events in terms of a binary value of *ongoing* or *ended*. This factor takes into account during-reporting continuity of events. Our survey included self-reported entries of the status of event, and for social media, we manually identified the temporal status by going through the life event disclosure posts[1].

**Event Type.** As introduced earlier, our datasets (both social media and surveys) group the life events into six broad categories of School, Health, Personal, Financial, Work, and Local. While the self-reported survey data was already annotated with these categories by the participants, the social media data life event expressions were annotated by our annotation approach and codebook[1]. We use the categorical variable of life event type as covariates in our analyses. Besides, although our data contains labels of finer categories of life events (e.g., *vacation, health loss, bad weather, child birth*, etc.), the number of records per event is plausibly not significant for statistical power, and may lead to inconclusive or misleading results [30] In addition, theoretically an individual only experiences a limited number of life events per year [55, 98], so it would be impractical to include all possible life events without a significantly larger sample size than what we have. We validate this hypothesis by conducting a $\chi^2$-square test, which reveals $\chi^2$ = NaN and p = NA, suggesting not enough observations per finer categories of life event.

[1]The Supplementary Material provides our detailed codebook, and the codebooks to annotate intimacy, scope, and status.

**Table 3: Brief Description of Strategies and Considerations for Identifying Life Event Disclosures on Social Media.**

| What constitutes a life event disclosure? |
|---|
| **Present events with potentially significant impact in the future.** |
| We coded posts as life events disclosing an event in the present which is significant enough that the individual would be able to recall it in a few years, or if the event in disclosure could potentially leave a significant emotional impact in the future. For example, *"Horrible day for travel. Two canceled flights and 2 delays. Sharing the sights from this week while I wait to get home."* |
| **Past events with significant emotional impact in the present.** |
| We also found self-disclosures about events from the past. Recalling these events conveys the significance of the event in the individual's life and leaves emotional impact. Therefore, for events that occurred a while ago, if they have a big enough emotional impact even in the present, these posts would be identified as a life event, e.g., recollecting the death of someone close often results in grief in the present [118], such as, *"When you are looking for one child's birth certificate and you find the other's with her death certificate.. 33 days and you would be 16.."* |
| **Using the post wording.** |
| Wherever applicable, in cases of close tie in assigning a post with a life event category, we prioritized the wording in the post. We considered that the individual's self-description of an event is less biased and closer to self-perceived life event type. For example, when deciding between *trip* and *vacation*, if the post explicitly used either of the two words, we assigned the same life event category. For example, we assigned *trip* for *"For my recent business trip I flew Delta. I'm giving them 4 stars. They have on-demand in-flight movies and I got to watch Black Panther."* |
| **Underlying reason of an event.** |
| As above, when multiple categories could fit a post, we prioritized the one that seemed to be the underlying cause. Sometimes, other posts around the same date provided more context to make these decisions. For example, in the following post, although both *vacation* and *positive relationship* could be appropriate, *positive relationship* (anniversary) was the more underlying cause (also consistent with the individual's other posts around the same date), *"What a beautiful weekend celebrating our 10th Anniversary! So thankful for getting away to enjoy time together as husband and wife <3."* |
| **Disclosing multiple life events.** |
| Some posts may disclose multiple life events, some of which may also be continuous or ongoing events. For example, an ongoing vacation may include a birthday party, or a post about wedding planning may also talk about other investments, e.g., *"Going to start selling a small selection of simple car [..] Trying to make some money on the side for wedding and honeymoon, and my medications. Also gotta pay this damn hospital bill now."* |
| **Continuous Life Events** |
| Life events disclosures on social media may not necessarily be about discrete or one-off events, but could also be a continuous process. The availability of longitudinal data also enabled us to identify events lasting for a time period, e.g., *start*, *during*, and *end* of a vacation. Continuous events can consist of 1) a series of posts which together build a continuous event, 2) other posts providing context about a seemingly vague post at hand, and 3) a single post describing a continuous event. These are not necessarily exclusive and can co-occur. For example, a post describing a *"view"* or a *"beautiful city"* may seem vague, however, posts around the same date provided context that these are during-vacation activities. Again, a continuous life event can include related or unrelated life events within that period. |
| **Additional Life Events Categories** |
| As noted before, while annotating social media life event disclosures, we also included some form of open coding in our approach. This allowed us to include new categories, which might not directly be present in the PERI scale. For example, we added a new category of *Voted* for a post, *"I voted".* |
| **What does not constitute a life event disclosure?** |
| **Vague Post.** |
| Exclude if the posts is too vague to make a deduction of a life event, e.g., *"Waited for this FOR FOREVER!!!!!!!!."* |
| **Joke or Entertainment Media related.** |
| We found cases where a post did mention a life event, or keywords related to life events, but there were explicit expressions of these to be a joke, or a description about an event in a movie, TV show, video game etc, for example, *"The end... he died lol!"* |
| **Past events, but no significant emotional impact in the present.** |
| We found posts that described events or self-experiences from the past, but the person does not seem to be significantly affected in the present. An example post excluded based on this criteria includes, *"The meals, and especially the Blue Mountain coffee, were the best in Jamaica."* |
| **General shares or global events.** |
| Posts that consisted third-person or generic information (without any personal reference) based sharing were excluded to be considered as life events. For instance, an example post on political topic that was excluded, *"Retweeted University Department:. In four years as a student at University, Name had seven internships.[..] The experiences helped her decide what she wants in a career [..]"* |

*Baseline Attributes.* Social media and self-reported surveys are fundamentally two different data modalities, and it is important to control our models on an individual's baseline behavior on these modalities. Essentially, for each individual, we compute four baseline attributes — *social media baseline attributes* include, 1) total number of posts and 2) average length of post per individual, and *survey baseline attributes* include, 3) total number of responses and 4) average significance self-reported in each response. These baseline attributes go in as covariates in our models.

*4.3.2 Tests and Models.* We now explain our statistical models. We first obtain a union of all the life events recorded on social media and on survey as our total dataset ($D_T$). Then, we conduct a One-way Multivariate Analysis of Variance (MANOVA) tests on the combination of dependent variables of social media self-disclosure and survey self-report to the set of theory-driven covariates explained above. A statistical significance in MANOVA would reveal the importance of each covariate in explaining life event reportage on either or both of social media and surveys.

Next, to understand the direction of the factors in their associate with life event disclosure, we conduct two kinds of analyses drawn on nested logistic regression models — one on $D_T$ and the other on a subset, $D_S$ consisting of events recorded in one of the two modalities. This would allow us to examine the intricacies of each factor and their signed (positive or negative) importance in explaining reportage. We describe the two analyses below:

- *Convergence:* The first analysis studies whether a life event is likely to be recorded in *both social media and survey* modalities. On $D_T$, we build a binary logistic regression model that uses dependent variable as a binarized value based on the occurrence on both modalities, i.e., if the event is logged on both modalities, it is labeled as 1, otherwise 0. This model is referred to as $\mathbf{Model_1}$.
- *Divergence:* The second analysis is conducted on $D_S$, among life event records which are *not* recorded on both the modalities — what is the likelihood of it to be self-disclosed on social media versus self-reported on survey. This logistic regression model uses as dependent variable the binarized

value based on the occurrence on either of the modalities. That is, given an individual's life event log which does not occur at both modalities, it is labeled as 1 when self-disclosed online, and labeled as 0 when self-reported on survey. We refer this model as $\textbf{Model}_2$.

# 5 RESULTS

## 5.1 Distributions of Life Events

We present the distribution of life events reportage on both modalities by number of individuals in Figure 1a and Figure 1b. First, we note the heavy skew at $x$=0 for social media disclosures, which does not exist for survey self-reports — a key difference in the characteristic of the two data modalities. Out of the 14,359 Facebook posts, only 14% (2,031) express life events as per our annotation. In contrast, the survey is a dedicated effort directly asking the participants to log life events, so, 100% of its responses correspond to some form of self-perceived notion of life event per individual.

Next, Figure 1c shows the category-wise distribution of life events in the two modalities. Both the datasets show a prevalence of Personal life events — 39.5% among all survey responses, and a high 70.4% among all online disclosures. Interestingly, Work, which is significantly logged in survey self-reports (32.5%), appears low on social media (5.7%). Health events are recorded comparably on both surveys (9.5%) and on social media (8.3%).

Table 4 presents the top life events recorded on the two modalities. We find *vacation* scores the highest on both. In fact, *vacations* and *trips* occur more commonly across individuals as opposed to the rarity and uniqueness of other events. Our data suggests that Facebook's design and perceived use-case may facilitate individuals to post prevalently about *vacation* and *trip* events. Again, these events are often recorded on calendars, which may guide individuals to report these events in the post-participation life events survey.

Table 4 also explains the significant occurrence of other categories in the self-reported survey data including, Work-related *performance review, promotions, heavy work*, and *job switches*, none of which are disclosed significantly on social media. Rather, the only Work categories frequently disclosed online are *good worklife* and *work success* — both of which bear a positivity in valence. This may indicate that people are not comfortable about sharing work-related negativity on social media due to concerns of employer surveillance [51]. Another interesting contrast includes that *health loss* appears as a top event self-reported in surveys, whereas *health gain* occurs in those disclosed online. These observations suggest an inclination towards disclosing positive events on social media, which may associate with perceived self-presentation and social desirability of individuals on a public platform (social media) [71].

We note the difference in labeling life events in the two modalities (self-perceived vs. inferred). This distinction may indirectly explain our observation that our annotation scheme identified increased social activities (e.g., celebrations, gatherings) as "life events", which might not be self-perceived the same way to be recalled during a survey that happened after a period of time. In contrast, *death in family* and *child birth* commonly occur in the top life events on both modalities. These events are known to bear both short-term as well as long-term effect on one individual's life [42].

## 5.2 Language of Life Event Disclosures

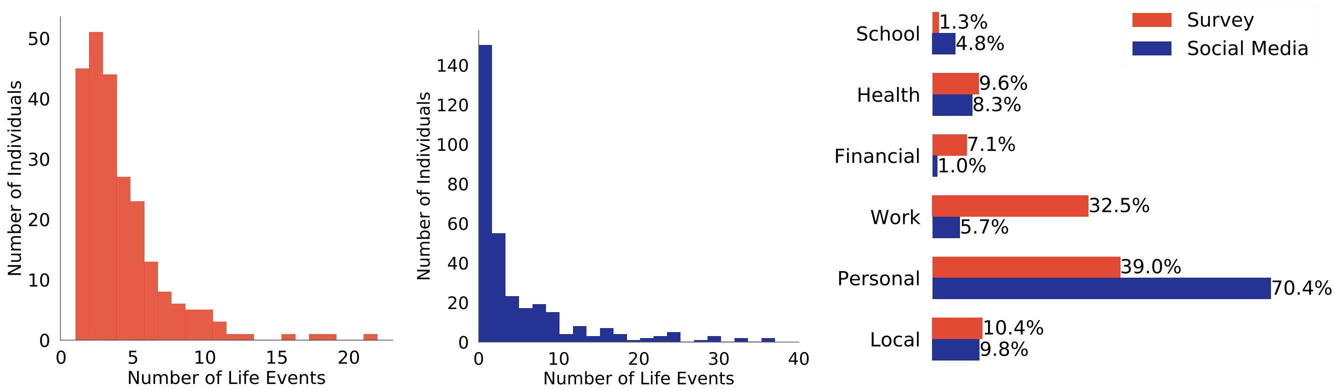
We are now interested to understand how individuals describe life event occurrences on their Facebook timelines. We investigate relationships between social media posts that were temporally similar to life events self-reported on surveys. In particular, for each individual, we look for events that were overlapping on the two modalities or occurred less than 7 days from each other. We aim to qualitatively determine what relationship, if any, there is between the reportage of life events on these modalities. After identifying pairs of potentially overlapping events from each modality, we compare and code the similarities and differences in linguistic descriptions of the events from the two modalities. Then, based on our codes, we conduct a thematic analysis to gradually coalesce the codes into themes of associating online disclosures and survey reported life event descriptions. We list some notable themes from our observations below.

**Emotional and Expressive Content.** Social media posts are more likely to bear an emotional tone about events. We find several occurrences for events such as adoption of pet and child birth, *"Name was born today. She was 8lbs 5oz and 21 inches long. We love her so much and are very thankful that she is happy and healthy! Thanks for all of the prayers!!"*. Similarly, social media posts also contain greater and richer detail about the event, for example, someone whose self-report survey entry only recorded a vacation, had posted on their social media about their vacation and positive relationship event, *"Best date night with my husband! Love you to the moon and back dear husband #wefishtogether."*

**Co-occurring and Related Events.** Sometimes the social media post can reflect a co-occurring and related event in someone's life. For example, an individual who self-reported on the survey to be on a vacation on certain dates, posted about a family meetup during those dates, *"Had the joy and privilege of seeing my niece dance in the ballet Sleeping Beauty today...also got to spend time with some people dear to my heart."*, here vacation and family meetup co-occur. Another individual, who changed jobs, posted about their move to a new city, *"Just rolled into California. Quite some driving but an easy roll into SF tomorrow."*

**Followup or Cause-Effect Related Events.** We observe instances where one life event may have triggered or caused a separate life event about which the individual posted on social media. For example, an individual who reported to be assaulted on a particular day, followed up with a Facebook post on *"I'm moving."* We also observe the opposite instance when an individual who self-reported about a bereavement leave at workplace on survey, had self-disclosed about the death of a family member a day prior to the reported date, *"This guy will be missed. Wish we had more time together [..]."*

**Co-occurring but Likely Unrelated Events.** Interestingly, we also observe instances of events that co-occur but are likely unrelated to each other. For example, an individual who self-reported on the survey having trouble with their boss at workplace, self-disclosed about their pet on social media, *"Help me find my foster pup a forever home! He is the sweetest and needs a great home asap [..]"* Again, another individual who self-reported on the survey about the death of a pet, had posted about a family reunion during the same time on social media, *"A family reunion time."*




**(a) Survey Self-reports** **(b) Social Media Disclosures** **(c) Category-wise Distribution**

**Figure 1: Distribution of data by life events (a) in self-reported survey data, (b) in social media data, (c) per category (percentage values are on all the life events reported within each dataset).**

**Table 4: Top life event recorded in survey self-reports and social media self-disclosures.**

| Survey Self-reports | | |
|---|---|---|
| **Life Event** | **Type** | **Count** |
| Vacation | Personal | 182 |
| Performance Review | Work | 117 |
| Bad Weather | Local | 88 |
| Health Loss | Health | 88 |
| Promoted | Work | 53 |
| Positive Job Switch | Work | 45 |
| Heavy Work | Work | 44 |
| Got Bonus | Work | 44 |
| Neutral Job Switch | Work | 42 |
| Trip | Personal | 40 |
| Installment Purchase | Financial | 36 |
| Child Birth | Personal | 33 |
| Death in Family | Personal | 28 |
| Positive Move | Personal | 28 |
| Financial Gain/Loss | Financial | 27 |

| Social Media Self-disclosures | | |
|---|---|---|
| **Life Event** | **Type** | **Count** |
| Vacation | Personal | 485 |
| Trip | Personal | 227 |
| Increased Social Activity | Personal | 142 |
| Family Meetups | Personal | 106 |
| Positive Relationship | Personal | 85 |
| Health Gain | Health | 69 |
| New Hobby | Personal | 67 |
| Positive Move | Personal | 56 |
| Death in Family | Personal | 45 |
| Back to School | School | 42 |
| Work Success | Work | 40 |
| Remodeled Home | Personal | 34 |
| Good Worklife | Work | 34 |
| Injury | Health | 34 |
| Child Birth | Health | 29 |

**Negative Stands Out in Recall.** We find instances where a negative event within a span of events outweighs the rest, and it is the only event reported in the survey (which happens later). In contrast, the social media data archives events from the past but were presumably recorded in-the-present. For example, in one instance, an individual posted about their ongoing vacation on social media, however, in the survey they only logged about a breakup on those dates. On another instance, an individual's social media data revealed them enjoying a vacation with friends, however they only self-reported a car-crash that might have happened then.

## 5.3 Factors Explaining Life Event Reportage

*5.3.1 Importance of Covariates in Reportage.* First, we examine the importance of our considered individual-centric and event-centric covariates in understanding people's disclosure of life events. For this, we conduct MANOVA tests as described in the previous section, with respect to the Pillai–Bartlett trace, which is considered to be robust and not strongly linked to normality assumptions the data distribution [111]. Table 5 summarizes the MANOVA statistics, where the F-statistic quantifies the association of the covariate with the dependent variables, and larger values indicate greater statistical importance. We next compare the $F$-statistic and significance across the covariates. Among the individual attributes, agreeableness (F=106.63) shows the greatest association, closely followed by gender (F=100.34). Among event attributes, status (F=988.62) and significance (F=592.16) show the greatest association, followed by anticipation (F=120.23) and valence (F= 85.43). The statistical significance shown by all variables (except anxiety) empirically validates our choice of the theory-driven variables we consider.

*5.3.2 Convergence: Reportage of Events on Both Social Media and Survey.* $\text{Model}_1$ examines the factors associated with life events reportage on *both of* against on *one of* the modalities (ref: Table 6). $\text{Model}_1$ shows a McFadden's pseudo $R^2$=0.18, $\chi^2(34)$=408.98 and $p < 0.001$, suggesting that the model is significantly better than an empty model. For a covariate $x$ showing a standardized coefficient estimate of $e$ with statistical significance, we interpret that a change

**Table 5: Multi-variate Analysis of Variance (MANOVA) results, * $p$<0.05, ** $p$<0.01, *** $p$<0.001.**

| Demographic/Trait | Pillai | F | p | Event Attribute | Pillai | F | p |
|---|---|---|---|---|---|---|---|
| Age | 0.036 | 47.01 | *** | Valence | 0.063 | 85.43 | *** |
| Gender | 0.072 | 100.34 | *** | Significance | 0.317 | 592.16 | *** |
| Born in US | 0.004 | 4.87 | ** | Recency | 0.044 | 59.81 | *** |
| Education | 0.051 | 16.73 | *** | Anticipation | 0.086 | 120.23 | *** |
| Shipley: Abstraction | 0.057 | 76.61 | *** | Intimacy | 0.002 | 3.11 | * |
| Shipley: Vocabulary | 0.033 | 42.97 | *** | Scope | 0.005 | 1.67 | * |
| Personality: Openness | 0.002 | 2.56 | * | Status | 0.516 | 988.62 | *** |
| Personality: Conscientiousness | 0.022 | 28.21 | *** | Type | 0.183 | 51.31 | *** |
| Personality: Extraversion | 0.003 | 4.21 | * | | | | |
| Personality: Agreeableness | 0.077 | 106.63 | *** | **Baseline Attribute** | **Pillai** | **F** | **p** |
| Personality: Neuroticism | 0.030 | 39.64 | *** | SM: Num. Posts | 0.088 | 121.00 | *** |
| Positive Affect | 0.003 | 3.73 | * | SM: Avg. Post Length | 0.003 | 4.04 | * |
| Negative Affect | 0.005 | 7.03 | *** | SR: Num. Records | 0.108 | 152.10 | *** |
| STAI: Anxiety | 0.001 | 1.56 | | SR: Avg. Significance | 0.019 | 25.47 | *** |
| PSQI: Healthy Sleep Quality | 0.011 | 14.45 | ** | | | | |

in one unit of standard deviation likely results in $e$ standard deviation change in the log odds of the dependent variable. In the case of $\texttt{Model}_1$, a positive coefficient indicates a propensity to reporting a life event on both modalities, and a negative coefficient indicates a propensity to report on one of the modalities.

Among demographics, we find that the likelihood to report on both modalities lowers as age increases. Similarly, males are less likely to report on both. This aligns with prior work [11] that males tend to self-disclose lesser than females on certain personal life events. Among traits, crystallized cognitive ability shows a significant positive coefficient. This is plausibly related to the notion that greater cognitive ability is known to drive the ability to distinguish positivity and negativity of situations to accordingly structure emotional expressiveness [127]. In personality traits, conscientiousness and agreeableness are significant, each showing opposite association — conscientiousness negatively associates whereas agreeableness positively associates with the likelihood to report on both modalities. Conscientiousness characterizes one's thoroughness [143] — a significance may be associated with individuals being methodical in delineating what they want to disclose on social media. On the other hand, agreeableness characterizes warmth and friendliness — an individual scoring high on agreeableness likely "gets along well" with others [143, 153]. This plausibly relates to people knowing their online audience better, and experiencing low inhibition to report on both modalities. Affect and wellbeing traits show weak relationships, and interestingly positive and negative affect exhibit opposite directions — higher positive affect explains lower reportage, whereas higher negative affect explains greater reportage on both modalities.

Among event attributes, we find event significance bears a strong negative coefficient (e=−0.33) indicating that significant events are less likely to be reported on both modalities. Anticipated events are likely to be reported on both (e=0.16); these events bear some form of planning or apriori awareness (e.g., child birth), and people may not only disclose them online, but also recall and report them in retrospective survey. In contrast, unanticipated events plausibly relate to emergency circumstances, and people may deprioritize an immediate online disclosure. These could also be short-term events (e.g., a positive relationship act) which may be disclosed on social media in-the-present, but may not remain in one's long-term memory to be self-reported in a survey which happened after a while. Among event types, Health and School events have propensity to be recorded on both social media and surveys, whereas, Work and Financial events are unlikely to be recorded on both modalities.

Finally, we also note the statistical significance of controlling for baseline behavior of individuals. Recording on both modalities shows a positive association with individuals who typically have more social media posts, more survey records, and whose baseline average significance of self-reported life events on survey is higher. However, average length of social media posts shows no statistical significance with respect to recording behavior.

*5.3.3 Divergence: Reportage of Events on Social Media Versus on Survey.* $\texttt{Model}_2$ examines the factors that associate with reporting life events on *either of* the two modalities (ref: Table 7). $\texttt{Model}_2$ shows a McFadden's pseudo $R^2$=0.77, $\chi^2(34)$=1785.83 with p<0.001, i.e., the model is significantly better than an empty model. Here positive coefficients suggest a propensity to record online, and negative suggests a propensity to report on survey (and not online).

Among individual-centric attributes, males (e=−0.38) show a lower likelihood to self-disclose online. This observation somewhat supports prior work that found men to show lower online self-disclosure than women [141]. We notice a strong association with agreeableness (e=0.73) — indicating that individuals with greater agreeableness have a likelihood to self-disclose life events on social media. Similarly, extraversion shows a positive coefficient (e=0.13). Extraversion characterizes one's outgoing, talkative, and energetic behavior [153], and this trait is known to associate with greater expressiveness and disclosure [113, 126]. We also see a weak negative significance for negative affect (e=−0.05), indicating that individuals scoring high on negative affect are less likely to disclose on social media, which could be associated with privacy and audience perceptions as noted in prior work [34, 96].

Among event-centric attributes, we find that valence (e=0.45) and anticipation (e=0.45) bear positive coefficients. This suggests that individuals tend to mostly disclose events on social media

**Table 6: Model$_1$ (Convergence): Coefficients of linear regression of relevant covariates as independent variables and reporting on both modalities as dependent variable, * $p$<0.05, ** $p$<0.01, *** $p$<0.001. Bar length is proportional to the magnitude of coefficient; for significant rows, orange bars (positive coefficients) indicate a propensity to record on both social media and survey, whereas teal bars (negative coefficients) bars indicate a propensity to record on one of the modalities.**

| Demographic/Trait | Std. Coeff. | p |
|---|---|---|
| Age | −0.03 | ** |
| Gender: Male | −0.41 | *** |
| Born in US: Yes | 0.41 | |
| Education: H. School | 1.57 | *** |
| Education: College | 1.33 | ** |
| Education: Grad School | 1.78 | ** |
| Education: Doctoral | 1.31 | * |
| Shipley: Abstraction | −0.03 | |
| Shipley: Vocabulary | 0.05 | ** |
| Personality: Openness | −0.24 | |
| Personality: Conscientiousness | −0.25 | * |
| Personality: Extraversion | 0.04 | |
| Personality: Agreeableness | 0.49 | *** |
| Personality: Neuroticism | 0.06 | |
| Positive Affect | −0.04 | * |
| Negative Affect | 0.06 | *** |
| Stai: Anxiety | −0.03 | * |
| PSQI: Healthy Sleep Quality | 0.02 | |

| Event Attribute | Std. Coeff. | p |
|---|---|---|
| Valence: Positive | 0.24 | |
| Significance | −0.33 | *** |
| Recency | −0.24 | |
| Ancptn.: Anticipated | 0.16 | * |
| Intimacy | 0.08 | |
| Scope | −0.51 | ** |
| Status: Ongoing | 1.08 | *** |
| Type: Health | 0.82 | ** |
| Type: School | 0.54 | * |
| Type: Work | −0.61 | * |
| Type: Local | −0.60 | ** |
| Type: Financial | −0.49 | ** |

| Baseline Attribute | Std. Coeff. | p |
|---|---|---|
| SM: Num. Posts | 0.48 | *** |
| SM: Avg. Post Length | 0.50 | |
| SR: Num. Records | 0.33 | ** |
| SR: Avg. Significance | 0.20 | ** |

AIC = 2047.40, Deg. Freedom= 33, Log-likelihood = −988.71, $\chi^2$= 408.98, McFadden's Pseudo $R^2$ = 0.18, p < 0.001 ***

**Table 7: Model$_2$ (Divergence): Coefficients of linear regression of relevant covariates as independent variables and reporting on either modality (1 for online/social media and 0 for survey) as dependent variable, * $p$<0.05, ** $p$<0.01, *** $p$<0.001. Bar length is proportional to the magnitude of coefficient; for significant rows, blue bars (positive coefficient) indicate a propensity to record only on social media, whereas red bars (negative coefficient) indicate a propensity to record only on survey.**

| Demographic/Trait | Std. Coeff. | p |
|---|---|---|
| Age | 0.04 | *** |
| Gender: Male | −0.38 | * |
| Born in US: Yes | −0.75 | |
| Education: H. School | 0.43 | |
| Education: College | 0.43 | |
| Education: Grad School | 0.47 | * |
| Education: Doctoral | 0.48 | |
| Shipley: Abstraction | −0.12 | *** |
| Shipley: Vocabulary | −0.05 | * |
| Personality: Openness | 0.18 | |
| Personality: Conscientiousness | −0.04 | |
| Personality: Extraversion | 0.13 | * |
| Personality: Agreeableness | 0.73 | *** |
| Personality: Neuroticism | −0.11 | |
| Positive Affect | 0.03 | |
| Negative Affect | −0.05 | * |
| Stai: Anxiety | 0.04 | |
| PSQI: Healthy Sleep Quality | −0.04 | |

| Event Attribute | Std. Coeff. | p |
|---|---|---|
| Valence: Positive | 0.45 | *** |
| Significance | −1.40 | *** |
| Recency | −3.56 | *** |
| Anticipated | 0.45 | * |
| Intimacy | −0.75 | ** |
| Scope | −0.93 | *** |
| Status: Ongoing | 3.62 | *** |
| Type: Health | −0.98 | |
| Type: School | 0.18 | |
| Type: Work | −1.18 | *** |
| Type: Local | −1.11 | * |
| Type: Financial | −2.90 | *** |

| Baseline Attribute | Std. Coeff. | p |
|---|---|---|
| SM: Num. Posts | 0.90 | *** |
| SM: Avg. Posts Length | −1.59 | |
| SR: Num. Records | 0.49 | *** |
| SR: Avg. Significance | −1.57 | *** |

AIC = 628.26, Deg. Freedom= 34, Log-likelihood = −279.13, $\chi^2$= 1785.83, McFadden's Pseudo $R^2$ = 0.77, p < 0.001 ***

that are positive and/or that are anticipated. On the other hand, both significance (e=−1.40) and recency (e=−2.90) bear strong negative coefficients. This supports prior research regarding the bias of self-reported surveys due to retrospective recall and significance of events [154]. Also, intimacy (e=-0.75) and scope (e=-1.03) bear negative coefficients, likely related to the public-facing nature of social media and people's self-presentation. Unsurprisingly, social media disclosures are also more skewed towards ongoing events because they enable in-the-present sharing, unlike surveys that elicit retrospective recollection.

Among life event types, Financial, Work, and Local events bear low likelihood to be disclosed online. People may not be comfortable about sharing their financial gain or loss events publicly on social media, or they may not share work-related events, especially if they have concerns around context collapse [96]. In contrast, School events may not be deemed that private, and people may be comfortable sharing about school-related success and milestones.

Finally, among baseline attributes, number of social media posts positively associates with life event disclosures on social media. Again, number of survey records also positively associates with social media disclosure. However, individuals who reported higher significance of events on average tend to post lower on social media — this could relate with people’s baseline perceptions of event significance and social media disclosures.

# 6 DISCUSSION

## 6.1 Theoretical Underpinnings & Implications

We sought to examine how/when people tend to disclose life events on social media, and the attributes of individuals who choose to disclose versus not. This section first discusses the theoretical underpinnings of our work, drawing on a host of theories and conceptualizations in social computing and HCI. To situate the validity of the disparities between disclosure and non-disclosure, we compared and contrasted social media disclosures with survey self-reports of life events — the latter being the gold standard in capturing life events. Accordingly, we also discuss how some of these differences are rooted in the differences in the two modalities in their context of use and available affordances.

*6.1.1 The Role of Audience and Norms.* Social desirability is a known bias in surveys and in face-to-face offline settings [60]. Our work reinforces prior evidence that this factor could potentially modulate social media disclosures as well [104]. We found instances when individuals were comfortable to disclose positive or anticipated events on social media that were not remembered during the survey. This may indicate a varied set of self-presentation goals propelled by the positivity bias in normative Facebook use [22], or a desire for selective “performance” as per Goffman’s “frontstage/ backstage” metaphor for impression management and social roles enactment [54], or for exhibitionism [71], or for receiving instant or short-term social approval and gratification [165].

These disclosures may also stem from a need to maintain and bridge social capital around transitory or minor happenings in one’s life, where sharing certain milestones, such an imminent wedding, or leaving/starting a job has become customary — a recent survey found that people “prefer sharing life’s milestones with their social network than in person” [18]. In fact, sharing life milestones on social media may not only revive dormant social connections, and simultaneously elicit responses or communication from an individual’s passive or weak ties [145], but also enhance the emotional tone and impact of the event [27]. Finally, positive and anticipated life event disclosures may also be attributed to the “desire to use online social media as a way for archiving life experiences and reflecting on identities,” especially if the events are associated with liminality [57]. Taken together, our findings shine a light on how the underlying norms of a social media platform, as well as its relationship to social desirability and impression management, may impact the semantics of a life event from an individual’s perspective, and the decision surrounding its online disclosure.

Complementarily, as societal norms motivate people to behave in particular ways [135], a social media platform’s norms may encourage certain disclosures as well as impose certain expectations that discourage people from sharing specific life events. Drawing on the literature on social comparison in social media [22, 117], people may not disclose very sensitive events such as an extra-marital relationship, a family conflict, or pregnancy loss for fear of social disenfranchisement, stigma, or shame, [6]. In fact, our study found that individuals withheld disclosing work-related and finance-related events on social media despite their occurrences per self-reports on the survey. Building upon the Disclosure-Decision Making framework proposed by Andalibi [6], we conjecture these decisions may be driven by people’s specific imagined or actual audiences [96] including their mental representations [109], wherein, due to concerns of context collapse [96], conflicting social spheres [16], surveillance [51], or the (semi-) public nature of the platforms [71], certain life events may be deemed less appropriate or share-worthy compared to others. Moreover, we found a lower likelihood of disclosing particularly intimate events or events too personal in their scope on social media. The design of the Facebook platform may in itself be a key factor driving self-regulatory decisions of non-disclosures [39]. Facebook particularly does not enable anonymity, a factor known to be facilitating intimate content sharing [94]. With an emphasis on “integrity and authenticity” as a community standard on the platform[2], other known disclosure risk mitigation strategies such as switching communication channels [61], using multiple accounts [161], or sharing incorrect information [81], may not apply for life event disclosures on Facebook.

Summarily, we draw upon Newman et al.’s [109] observations about sharing sensitive information on Facebook, that people carefully navigate the tension between sharing vulnerability, needs, and health status information and the desire to convey positive images of themselves. We note an apparent dichotomy that the same factors which encourage disclosure on Facebook (e.g., real identity, online and offline friendship networks, closed/known audience) for some instances (e.g., wedding) may also likely inhibit disclosure for some other instances (e.g., family conflict). Our work therefore emphasizes a need to understand the interplay between audience and norms of a life event reportage in the online context. This can be studied via the lens of the socio-technical gap [1] to understand the fundamental discrepancy in facilitation of socio-technical systems — what individuals disclose online and what they disclose offline, and how the technical design of the systems may encourage one set of practices or goals over the other [1].

*6.1.2 Contextual and Affordance Differences.* Our results showed a contrast between social media disclosures and survey self-reports, which elicits a discussion of the respective modalities’ affordances and context of use. We note that social media is naturalistic and largely recorded in-the-present unlike the survey which was retrospective and researcher-prompted; social media posting is also largely based on intrinsic motivation, whereas survey responses are

[2] https://www.facebook.com/communitystandards/integrity_authenticity

driven by extrinsic motivation (e.g., monetary incentive). That said, both require individuals' active effort to be recorded. Accordingly, we derive an interesting relationship with valence, significance and recency, and the ongoing nature of the life events — event attributes along which the reportage significantly differed (Table 7).

To start off, as discussed above, audience and impression management norms may make social media platforms to be less predisposed to sharing negative life events. However, why did our participants feel comfortable sharing negative life events with an audience of researchers? Compared to the social media audience that likely consists of strong and weak ties spanning online and offline interactions, researchers were strangers to the participants and comprised a smaller and likely perceived to be a more private audience than social media audience. These factors may have facilitated self-reporting of negative life events, free from concerns of stigma, social acceptance, or negative self-image.

Second, our findings support prior work that self-reported survey responses to likely be skewed to significant and recent events — significance and recency may cause disparities in emotional content, or salience, as these factors can change over time, especially after long time frames; emotional arousal may decay over time [33]. Extant literature lacks similar knowledge about online life events disclosures. Our work contributes to this knowledge that significance and recency negatively associate with social media disclosures. The immediacy of active attention needed for a significant event may explain the lower likelihood of online posting. For instance, during a health emergency, an individual may not actively record a social media post, as the situation may demand attention to other more immediate, important needs. Again, in specific circumstances, significance of an event could be hard to understand in-the-present but may be realized only after a period of time [42], e.g., a dinner outing with a friend that becomes memorable after the friend's sudden, unexpected demise. Evolving significance can also lead to a different impression in memory, such as a case in our study when an individual posted about a vacation (with their significant other) on social media, but only self-reported about a breakup in the survey. Presumably, when the vacation began and was shared on social media, it initiated positive feelings, but after it ended with a breakup, the negative event stuck in the individual's memory.

Third, ongoing events are more likely to be shared on social media versus a survey, and that might relate to the social affordances of social media such as private messaging or an ability to write on someone's timeline; e.g., an individual in the process of moving between two places may feel like they can gather help, support, and advice relating to the move, as the event unfolds in real-time. These social affordances were absent in the survey conducted in our study, since the audience constituted the researchers, and the participants were recounting about life events from the past.

Ultimately, both in-the-present and retrospective perception of an event may depend on an individual's coping process [41, 163]. While validated surveys can measure how an individual coped with a traumatic or stressful life event, social media data can provide a stream of in-the-present recordings, e.g., our dataset contained a series of posts explaining the logistics, stress and support related to hospitalization process of an individual's child (identified as a continuous category). Surveys may also cause priming effects [136] — if a participant is inquired about a stressful life experience, they may undergo a psychological stress by re-thinking about those experiences. Considering these differences, our work shows that additional factors relating to events and individuals are important drivers of disclosures (and non-disclosures). To this end, our study also extends prior investigations that have examined the factors behind disclosure and non-disclosure on social media alone [6], by asking questions around how individuals arrive at decisions regarding which life event to disclose on social media versus self-report on a survey, and how these decisions straddle the contexts of use and affordances of the two modalities.

## 6.2 Design Implications

As noted in Section 2, considerable HCI research has sought to design, develop, and adapt platforms around life events like childbirth [36], gender transition [65, 66], and pregnancy loss [7]. Going beyond instances of specific life events, our work reveals that people not only share varied life events on social media, but also engage in selective sharing of life events, controlling for individual differences and event attributes. Our research reveals, for the first time, a need to design for individuals and situations for both when disclosures do happen and when disclosures are withheld. Doing so necessitates closing the socio-technical gap per Ackerman [1].

*6.2.1 Designing for Disclosure.* We include two design implications here, based on our findings, one to scaffold the disclosure process itself, and a second to make platforms and their algorithms sensitive to disclosures once they happen.

Prior work reveals therapeutic and positive benefits of disclosure and expressive writing [13, 49], including benefits like finding an outlet for emotional release, self-acceptance, and solidarity with peers with similar experiences. Our work finds that despite the occurrences of negative life events, individuals may not always disclose these events on social media, perhaps because of concerns noted in Section 6.1. As also noted by Andalibi and Forte [7] and Ernala et al. [49], future research can therefore explore designing social media affordances that provide safe spaces for opening up for individuals with varied needs. This can include enabling individuals to create "trusted friend circles" based on various life event disclosures, e.g., a person may not be comfortable about sharing a work-related event but may be comfortable doing so with a set of trusted group of friends, therefore allowing targeted and staged disclosures [67, 160]. We found that users might be inhibited about disclosing negative or sensitive events. Users chose to not disclose certain events, despite Facebook providing audience control by design. To ease the process of recording an event privately or selectively, features may be included whose design and user experience are explicitly tailored to support the specific activity of recording life events, such as empowering users to define audiences and to limit the responses types about their life event, letting them take conversations to a different medium or outside of the platform, or having the provision of an expiration date on how long a life event may remain shared.

In addition, social media has shown promise as an intervention medium for crisis and wellbeing [29, 134]; we need to re-think alternative strategies for self-disclosures. For instance, to support individuals concerned about the public-facing nature of online platforms, tools may be built that emulate the benefits of personal blogging and journaling [31], to serve as a timestamped archive of one's thoughts and feelings around life events, empowering individuals to self-reflect traces of life. This can be a part of identity work or a part of memory work. We also found that disclosure behavior may reveal an individual's momentary and longitudinal behavior, such as some disclosures being associated with momentary affective states (e.g., grief and joy), and others with lasting changes (e.g., moving to a different place). Consequently, we suggest designing tools to provide supportive interventions around disclosures, including suggestions to rekindle interactions with social ties or recommending support communities.

On personal journaling, Facebook currently allows users to post and limit visibility to private. Some users send messages to themselves to record various events. However, none of these are by-design journaling interfaces. A recommendation could be an explicit private timeline space, where users can write private notes. Drawing motivation from smart journaling [47], such design can enable users to record life events, choose what to keep public and private, and also to toggle a private life event as public later in time. Further, platforms can consider designing with flexible anonymity, which can help break stereotyping or social expectations about social media posting of specific life events by particular demographics such as males and younger adults (as also seen in our study).

Next, as our Introduction notes, algorithmic content recommendation on social media is largely content and interests driven, showing personalized content based on individuals' interests and interactions with social ties. A lack of alignment of these recommendations with happenings in one's life, whether disclosed or undisclosed, can however have deep negative repercussions. We noted an anecdote when algorithmic curation of Facebook feed was "inadvertently cruel" because it were not sensitive to an individual's life event [102]. Therefore, like prior HCI work [8, 23], we argue that tailoring recommendations to be inclusive and attuned to disclosed life events can increase the value people derive from these platforms. Literature notes that positive content can potentially benefit individuals to feel better in positive times, whereas supportive content may enable to feel comforted during adverse times [8, 110]. Such uses of social media can be promoted by designing life event-inclusive and -aware recommendation algorithms and affordances.

*6.2.2 Designing for Non-Disclosure.* Our study reveals that a "one size fits all" approach to scaffold online life event disclosures may not work. It matters not only that certain individuals choose not to disclose, but also that each event is associated with unique characteristics and circumstances. In particular, although our study did not solicit feedback from participants about why they chose to disclose or not disclose a particular event, we did find certain demographic groups, such as males, older individuals, those low on agreeableness and extraversion personality traits less inclined to disclosing online. Essentially, from a therapeutic perspective, the perceived efficacy of social media platforms as online social spaces to disclose life events, may vary across individuals. Despite having a Facebook account and using Facebook for other purposes, individuals may resist or reject using the platform to share personal happenings, as an individual choice, social practice, or the event's temporality — a case for many of our participants. Scholars exploring technology non-use have found that disenchantment often stems from the perceived banality and inauthenticity of social interactions on social media platforms, particularly in contrast to offline communication [12]. Furthermore, some might feel socially disenfranchised to participate on a platform due to socio-institutional pressures, harassment, or social anxiety [115]. Because a disclosure might compromise an individual's social network's contextual integrity and the privacy expectations of other stakeholders of the life event [6], some of these factors behind non-use might play in our case as well. And yet, there were individuals who felt comfortable to self-report a life event on the survey, to a different social audience (of researchers), albeit smaller — indicating an implicit effort to weigh in the benefits and risks of disclosing life events on one modality versus another.

So how do we then design to accommodate the needs of these individuals with varying underlying decision-making processes around life event disclosures, and what would constitute an efficacious social media platform design for them? Given that our study reveals specific demographic differences among those who disclose and do not, how can design ensure that the groups who do not disclose are not marginalized?

Instead of designing only to encourage life event sharing on social media and risking "problematizing" the non-disclosers, we provide design suggestions drawing from scholars who have called for the role and perspective of the non-user to be recognized and valued [166]. First, platform designers need to account for social media non-use as a signal to modulate content recommendations. Essentially, design features may be built that allow individuals to curate or select what they would like to see and not see on the platform, depending on whatever their undisclosed current life event(s) might be. Second, drawing upon research on designing for technology non-use [12, 26, 122], platforms can accommodate alternative forms of participation for an individual, as a coping mechanism following an undisclosed life event, that does not involve being forced to deactivate or delete their social media account, or to stop social sharing and interaction altogether. For instance, individuals can switch platform settings to "no recommended content" and only visit parts of the site which they may feel are conducive to their current life circumstances. Broadly speaking, we draw from Baumer et al. [12], who noted that resistance to early telephone and electrical technology, particularly among rural populations, led producers to develop new designs and infrastructures better suited to rural life [89]. Similarly, we urge researchers and designers to make social media platforms life event-sensitive in a way that not only considers potential barriers preventing disclosures, but also provides agency in the decision-making processes behind non-disclosures.

## 6.3 Ethical Implications

Our work has ethical implications. While some of the motivations and implications of our work center around designing social media platforms that can customize content depending on individuals' life events, we note that personalization can function as a "double-edged sword" [116]. Pandit and Lewis argue that on one side, it can provide benefits through personalization and user profiling, but simultaneously can also raise several ethical and moral questions [116]. Despite the best of intentions of a platform and designers to provide personalized content, this can lead to expectation mismatches, and individuals may perceive intrusiveness and dissatisfaction about such algorithmic content curation without consent [50]. Further, identifying life event disclosures on social media can lead to other potential ethically questionable consequences such as targeted advertising [76], including compromised privacy, defying expectations, and damaging relationships — reminiscent of the case of the woman whose pregnancy was discovered by a supermarket chain without her knowledge [70]. That is, although personalizing ads around positive events (e.g., new home, wedding) may bear both business and individual advantages, the same around negative life events can not only exacerbate an individual's situation and wellbeing, but also can be deemed unethical and intrusive [92].

Furthermore, people's online disclosures of life events can be (mis)used to infer high-risk decision outcomes in one's offline life such as job, insurance coverage, financial support, or obtaining a property mortgage. At the other end of the spectrum, when people do not disclose their life events, it might prevent such misuse, but they may be disadvantaged in deriving the benefits that disclosing individuals might be able to derive from the platform, such as access to support, social capital, or social approval. From a social computing standpoint, both disclosures and non-disclosures of life events on social media can lead to forming new social conventions and norms on the platform with repercussions on an individual's life, e.g., research already notes the positivity bias on social media [22], and non-disclosure of negative events may make people feel worse when they experience a negative life event. Overall, these ethical complexities call for better understanding and guidelines regarding what platforms owners and decision makers can and should do with people's (non)-disclosures of life events, for what purpose, and the extent to which transparency is baked into these uses.

### 6.4 Limitations and Future Directions

Our work has some limitations, which suggest opportunities for future research. While we explored several factors related to life event disclosures on social media, one aspect that remains to be explored more concretely is the "why" question about people's self-disclosures. Our work can neither claim causality, and nor can it explain the causal directions (if any) between the factors and disclosure behavior. Future work can interview individuals to understand the causes of different behavior on social media and elsewhere, regarding disclosure (or no disclosure) of a life event.

Next, self-reports on a survey are based on an individual's subjective perception of interpreting life events. In contrast, labels on social media disclosures relied on the annotation scheme provided by our codebook, which essentially normalized the semantics of life events across all individuals' data — arguably less sensitive to subjective interpretation. However, this data can be prone to researcher bias, based on how our annotators read an individual's post, and the plausible *interpretation gap* in what the individual meant and what the annotators interpreted to (not) be a life event. Future work can consider to augment this study design where the codebook is adapted to each individual's subjective interpretation of life events, based on explicit feedback from them, in order to minimize the interpretation gap.

Moreover, we also note the lack of availability of real ground-truth data on the life events to corroborate the authenticity of social media life event disclosures. This issue is especially significant because although the PERI life event scale is a gold standard established in the literature [41], scholars have also noted biases that impact survey responses, such as, the fact that researchers are requesting personal data or the participants' own perceptions of the study for which they are sharing their responses [154]. Future research can investigate study designs to augment survey responses, such as using interviews or gathering data from participants' calendars or journals, in order to construct a more comprehensive picture of significant events in an individual's life.

Our findings are limited to a single data source, Facebook, and on those who chose to participate in the study, likely introducing self-selection bias. Each modality can have its own social conventions and expectations [117], contributing to an individual's self-disclosure on a particular topic in a particular way. Again, an individual is most likely active on multiple social media platforms for different purposes and audiences. Future work can extend this work to shed light on life events disclosure within and across multiple platforms, with participant consent like used here. For privacy reasons, our study does not include multimedia (e.g., photos) and private messages, these forms of data, again subject to participants' comfort levels, can contextualize the observations related to certain forms disclosures. Finally, our study is limited to examining only active participation on social media (posting on Facebook). We chose to exclude 242 participants who self-reported on survey but did not post on Facebook in the same period in our analysis, as this could have led to inconclusive information about their use of Facebook during study period, i.e., we cannot delineate if they were absolutely inactive on Facebook, or if they only passively participated (consumed) content on Facebook. If login/consumption data is available, future work can provide additional valuable insights on social media disclosures of life events.

## 7 CONCLUSION

This study examined how life events are recorded on social media, in terms of what is disclosed (or not), when, and by whom. We compared social media disclosures of life events on 256 participants' year-long Facebook dataset of 14K posts, against self-reported life event occurrences in this period. We defined and contributed a comprehensive codebook to identify online self-disclosures of life events. We examined what factors explain the deviation of online self-disclosed life events against self-reported life events. We built regression models by controlling for individual attributes such as demographics and intrinsic traits and event-centric attributes. We found that positive and anticipated events are more likely to be disclosed, whereas significant, recent, and intimate events are less likely to be disclosed on social media. Our observations suggested that all individuals might not disclose all life events on social media; however, what they disclose, provides complementary and richer information compared to what their self-reports reflect.

## ACKNOWLEDGMENTS

This research is supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800007. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the

U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. We thank the entire Tesserae team, Sarah Yoo, Yujia Gao, and Shrija Mishra for contributing to this work. We also thank Dong Whi Yoo, Qiaosi Wang, and the members of the Social Dynamics and Wellbeing Lab for feedback.

## REFERENCES


[1] Mark S Ackerman. 2000. The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction* 15, 2-3 (2000), 179–203.

[2] Daniel E Adkins, Victor Wang, Matthew E Dupre, Edwin JCG Van den Oord, and Glen H Elder Jr. 2009. Structure and stress: Trajectories of depressive symptoms across adolescence and young adulthood. *Social forces* 88, 1 (2009), 31–60.

[3] Icek Ajzen. 1985. From intentions to actions: A theory of planned behavior. In *Action control*. Springer, 11–39.

[4] Irwin Altman. 1975. The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding. (1975).

[5] Irwin Altman and Dalmas A Taylor. 1973. *Social penetration: The development of interpersonal relationships.* Holt, Rinehart & Winston.

[6] Nazanin Andalibi. 2020. Disclosure, Privacy, and Stigma on Social Media: Examining Non-Disclosure of Distressing Experiences. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–43.

[7] Nazanin Andalibi and Andrea Forte. 2018. Announcing pregnancy loss on Facebook: A decision-making framework for stigmatized disclosures on identified social network sites. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[8] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proc. CHI*.

[9] 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*. 1485–1500.

[10] Alan D Baddeley and Graham Hitch. 1993. The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition* 21, 2 (1993), 146–155.

[11] Mitchell K Bartholomew, Sarah J Schoppe-Sullivan, Michael Glassman, Claire M Kamp Dush, and Jason M Sullivan. 2012. New parents' Facebook use at the transition to parenthood. *Family relations* 61, 3 (2012), 455–469.

[12] Eric PS Baumer, Phil Adams, Vera D Khovanskaya, Tony C Liao, Madeline E Smith, Victoria Schwanda Sosik, and Kaiton Williams. 2013. Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*.

[13] Natalya N Bazarova and Yoon Hyung Choi. 2014. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication* 64, 4 (2014), 635–657.

[14] Michael S Bernstein, Eytan Bakshy, Moira Burke, and Brian Karrer. 2013. Quantifying the invisible audience in social networks. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 21–30.

[15] Jennifer L Bevan, Ruth Gomez, and Lisa Sparks. 2014. Disclosures about important life events on Facebook: Relationships with stress and quality of life. *Computers in Human Behavior* 39 (2014), 246–253.

[16] Jens Binder, Andrew Howes, and Alistair Sutcliffe. 2009. The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 965–974.

[17] boyd, danah and Ellison, Nicole B. 2007. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication* (2007).

[18] Elleen Brown. 2018. Americans prefer sharing life's milestones with their social network than in person. https://www.zdnet.com/article/americans-prefer-sharing-lifes-milestones-with-their-social-network-than-in-person/. Accessed: 2020-09-15.

[19] Jed R Brubaker, Gillian R Hayes, and Paul Dourish. 2013. Beyond the grave: Facebook as a site for the expansion of death and mourning. *The Information Society* 29, 3 (2013), 152–163.

[20] Jed R Brubaker, Funda Kivran-Swaine, Lee Taber, and Gillian R Hayes. 2012. Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media.. In *ICWSM*.

[21] Taina Bucher, Anne Helmond, et al. 2017. The affordances of social media platforms. *The SAGE handbook of social media* (2017), 233–253.

[22] Moira Burke, Justin Cheng, and Bethany de Gant. 2020. Social Comparison and Facebook: Feedback, Positivity, and Opportunities for Comparison. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

[23] 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*. 1419–1430.

[24] Ralph Catalano. 1979. *Health, behavior and the community: An ecological perspective*. Pergamon Press New York.

[25] Paulo R Cavalin, Luis G Moyano, and Pedro P Miranda. 2015. A multiple classifier system for classifying life events on social media. In *2015 IEEE international conference on Data mining workshop (ICDMW)*. IEEE, 1332–1335.

[26] Justin Cheng, Moira Burke, and Elena Goetz Davis. 2019. Understanding perceptions of problematic Facebook use: When people experience negative life impact and a lack of control. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[27] Mina Choi and Catalina L Toma. 2014. Social sharing through interpersonal media: Patterns and effects on emotional well-being. *Computers in Human Behavior* 36 (2014), 530–541.

[28] Smitashree Choudhury and Harith Alani. 2014. Personal life event detection from social media. (2014).

[29] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When personal tracking becomes social: Examining the use of Instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*.

[30] Jacob Cohen. 1992. Statistical power analysis. *Curr. Dir. Psychol. Sci.* (1992).

[31] Michael A Cohn, Matthias R Mehl, and James W Pennebaker. 2004. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science* 15, 10 (2004), 687–693.

[32] Bruce E Compas, Glen E Davis, and Carolyn J Forsythe. 1985. Characteristics of life events during adolescence. *American Journal of Community Psychology* 13, 6 (1985), 677–691.

[33] Tony J Cunningham, Charles R Crowell, Sara E Alger, Elizabeth A Kensinger, Michael A Villano, Stephen M Mattingly, and Jessica D Payne. 2014. Psychophysiological arousal at encoding leads to reduced reactivity but enhanced emotional memory following sleep. *Neurobiology of Learning and Memory* (2014).

[34] Sauvik Das and Adam Kramer. 2013. Self-censorship on Facebook. In *Proc. ICWSM*.

[35] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3267–3276.

[36] Munmun De Choudhury, Scott Counts, Eric Horvitz, and Aaron Hoff. 2014. Characterizing and Predicting Postpartum Depression from Facebook Data. In *Proc. CSCW*.

[37] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Eighth international AAAI conference on weblogs and social media*.

[38] Munmun De Choudhury and Michael Massimi. 2015. " She said yes"–Liminality and Engagement Announcements on Twitter. *iConference 2015 Proceedings* (2015).

[39] Michael A DeVito, Jeremy Birnholtz, and Jeffery T Hancock. 2017. Platforms, people, and perception: Using affordances to understand self-presentation on social media. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 740–754.

[40] Thomas Dickinson, Miriam Fernandez, Lisa A Thomas, Paul Mulholland, Pam Briggs, and Harith Alani. 2015. Identifying prominent life events on twitter. In *Proceedings of the 8th International Conference on Knowledge Capture*. 1–8.

[41] Barbara Snell Dohrenwend, Alexander R Askenasy, Larry Krasnoff, and Bruce P Dohrenwend. 1978. Exemplification of a method for scaling life events: The PERI Life Events Scale. *Journal of health and social behavior* (1978), 205–229.

[42] Barbara S Dohrenwend and Bruce P Dohrenwend. 1974. *Stressful life events: Their nature and effects.* John Wiley & Sons.

[43] Yuriko Doi, Masumi Minowa, Makoto Uchiyama, Masako Okawa, Keiko Kim, Kayo Shibui, and Yuichi Kamei. 2000. Psychometric assessment of subjective sleep quality using the Japanese version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects. *Psychiatry research* 97, 2-3 (2000), 165–172.

[44] Judith S Donath et al. 1999. Identity and deception in the virtual community. *Communities in cyberspace* 1996 (1999), 29–59.

[45] Robin IM Dunbar, Anna Marriott, and Neil DC Duncan. 1997. Human conversational behavior. *Human nature* 8, 3 (1997), 231–246.

[46] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. 2006. Managing impressions online: Self-presentation processes in the online dating environment. *Journal of computer-mediated communication* 11, 2 (2006), 415–441.

[47] Chris Elsden, Abigail C Durrant, and David S Kirk. 2016. It's Just My History Isn't It? Understanding smart journaling practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2819–2831.

[48] Sindhu Kiranmai Ernala, Tristan Labetoulle, Fred Bane, Michael L Birnbaum, Asra F Rizvi, John M Kane, and Munmun De Choudhury. 2018. Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. In *ICWSM*.

[49] Sindhu Kiranmai Ernala, Asra F Rizvi, Michael L Birnbaum, John M Kane, and Munmun De Choudhury. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *Proceedings of the ACM*

*on Human-Computer Interaction* 1, CSCW (2017), 1–27.
[50] Casey Fiesler and Nicholas Proferes. 2018. "Participant" perceptions of Twitter research ethics. *Social Media+ Society* (2018).
[51] Saby Ghoshray. 2013. Employer surveillance versus employee privacy: The new reality of social media and workplace privacy. *N. Ky. L. Rev.* 40 (2013), 593.
[52] Kimberly Glasgow, Clayton Fink, and Jordan L Boyd-Graber. 2014. " Our Grief is Unspeakable": Automatically Measuring the Community Impact of a Tragedy.. In *ICWSM*.
[53] Arthur M Glenberg, Margaret M Bradley, Thomas A Kraus, and Gary J Renzaglia. 1983. Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9, 2 (1983), 231.
[54] Erving Goffman. 1959. The presentation of self in everyday life. (1959).
[55] Evelyn L Goldberg and George W Comstock. 1980. Epidemiology of life events: Frequency in general populations. *American Journal of epidemiology* (1980).
[56] Matt J Gray, Brett T Litz, Julie L Hsu, and Thomas W Lombardo. 2004. Psychometric properties of the life events checklist. *Assessment* (2004).
[57] Samuel Greengard. 2012. Digitally possessed. *Commun. ACM* (2012).
[58] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Demographics of Social Media Users in 2016. pewinternet.org/2016/11/11/social-media-update-2016/. Accessed: 2017-02-12.
[59] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social media update 2016. *Pew Research Center* 11, 2 (2016), 1–18.
[60] Pamela Grimm. 2010. Social desirability bias. *Wiley international encyclopedia of marketing* (2010).
[61] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H Mhaidli, Kentaro Toyama, and Florian Schaub. 2018. Keeping a low profile? Technology, risk and privacy among undocumented immigrants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.
[62] Jamie Guillory and Jeffrey T Hancock. 2012. The effect of Linkedin on deception in resumes. *Cyberpsychology, Behavior, and Social Networking* (2012).
[63] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019. Understanding and measuring psychological stress using social media. In *Proc. ICWSM*.
[64] Oliver L Haimson, Nazanin Andalibi, Munmun De Choudhury, and Gillian R Hayes. 2018. Relationship breakup disclosures and media ideologies on Facebook. *New Media & Society* 20, 5 (2018), 1931–1952.
[65] Oliver L Haimson, Anne E Bowser, Edward F Melcer, and Elizabeth F Churchill. 2015. Online inspiration and exploration for identity reinvention. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3809–3818.
[66] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. 2015. Disclosure, stress, and support during gender transition on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1176–1190.
[67] Oliver L Haimson and Tiffany C Veinot. 2020. Coming Out to Doctors, Coming Out to "Everyone": Understanding the Average Sequence of Transgender Identity Disclosures Using Social Media Data. *Transgender Health* (2020).
[68] Kate L Harkness and Scott M Monroe. 2016. The assessment and measurement of adult life stress: Basic premises, operational principles, and design requirements. *Journal of Abnormal Psychology* 125, 5 (2016), 727.
[69] M Harris. 2005. Is journaling empowering? Students' perceptions of their reflective writing experience. *Health SA Gesondheid* 10, 2 (2005), 47–60.
[70] Kashmir Hill. 2012. How target figured out a teen girl was pregnant before her father did. *Forbes, Inc* (2012).
[71] Bernie Hogan. 2010. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* 30, 6 (2010), 377–386.
[72] Erin E Hollenbaugh and Amber L Ferris. 2014. Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. *Computers in Human Behavior* 30 (2014), 50–58.
[73] Thomas H Holmes and Richard H Rahe. 1967. The social readjustment rating scale. *Journal of psychosomatic research* (1967).
[74] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.
[75] CJ Hutto and E Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. ICWSM proceedings.
[76] Mark Irvine. 2018. Google Releases Life Events Targeting to Everyone! https://www.wordstream.com/blog/ws/2017/11/29/adwords-life-events-targeting. Accessed: 2020-09-08.
[77] Adam N Joinson. 2001. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology* 31, 2 (2001), 177–192.
[78] Adam N Joinson and Carina B Paine. 2007. Self-disclosure, privacy and the Internet. *The Oxford handbook of Internet psychology* (2007), 2374252.
[79] Sidney M Jourard. 1971. *Self-disclosure: An experimental analysis of the transparent self.* NewYork, NY: Wiley-Interscience.
[80] Sanjay Ram Kairam, Dan J Wang, and Jure Leskovec. 2012. The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining*. 673–682.
[81] Ruogu Kang, Stephanie Brown, and Sara Kiesler. 2013. Why do people seek anonymity on the internet? Informing policy and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2657–2666.
[82] Dilara Keküllüoglu, Walid Magdy, and Kami Vaniea. 2020. Analysing Privacy Leakage of Life Events on Twitter. In *ACM Web Science*.
[83] Kenneth S Kendler, John M Hettema, Frank Butera, Charles O Gardner, and Carol A Prescott. 2003. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. *Archives of general psychiatry* 60, 8 (2003), 789–796.
[84] Elizabeth A Kensinger and Suzanne Corkin. 2004. Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences* 101, 9 (2004), 3310–3315.
[85] Maryam Khodabakhsh, Mohsen Kahani, Ebrahim Bagheri, and Zeinab Noorian. 2018. Detecting life events from twitter based on temporal semantic features. *Knowledge-Based Systems* 148 (2018), 1–16.
[86] Emre Kıcıman, Scott Counts, and Melissa Gasser. 2018. Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use.. In *ICWSM*. 171–180.
[87] Eunice Kim, Jung-Ah Lee, Yongjun Sung, and Sejung Marina Choi. 2016. Predicting selfie-posting behavior on social networking sites: An extension of theory of planned behavior. *Computers in Human Behavior* 62 (2016), 116–123.
[88] Funda Kivran-Swaine, Sam Brody, Nicholas Diakopoulos, and Mor Naaman. 2012. Of joy and gender: emotional expression in online social networks. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*. 139–142.
[89] Ronald Kline. 2003. Resisting consumer technology in rural America: The telephone and electrification. *How users matter: The co-construction of users and technology* (2003), 51–66.
[90] MH Landis and Harold E Burtt. 1924. A Study of Conversations. *Journal of Comparative Psychology* 4, 1 (1924), 81.
[91] Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1997–2007.
[92] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In *Proc. ACM conference on Computer and communications security*.
[93] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing behavioral stages from social media data. In *CSCW*.
[94] Xiao Ma, Jeff Hancock, and Mor Naaman. 2016. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, 3857–3869.
[95] Gloria Mark, Mossaab Bagdouri, Leysia Palen, James Martin, Ban Al-Ani, and Kenneth Anderson. 2012. Blogs as a collective war diary. In *CSCW*. ACM, 37–46.
[96] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.
[97] Michael Massimi and Ronald M Baecker. 2010. A death in the family: opportunities for designing technologies for the bereaved. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1821–1830.
[98] Minoru Masuda and Thomas H Holmes. 1978. Life events: Perceptions and frequencies. *Psychosomatic medicine* 40, 3 (1978), 236–261.
[99] Stephen M Mattingly, Julie M Gregg, Pino Audia, Ayse Elvan Bayraktaroglu, Andrew T Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Sidney K DḾello, Anind K Dey, et al. 2019. The Tesserae Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. (2019).
[100] Andrew L Mendelson and Zizi Papacharissi. 2010. Look at us: Collective narcissism in college student Facebook photo galleries. *The networked self: Identity, community and culture on social network sites* 1974 (2010), 1–37.
[101] Dar Meshi, Loreen Mamerow, Evgeniya Kirilina, Carmen Morawetz, Daniel S Margulies, and Hauke R Heekeren. 2016. Sharing self-related information is associated with intrinsic functional connectivity of cortical midline brain regions. *Scientific Reports* 6, 1 (2016), 1–11.
[102] Eric Meyer. 2014. Inadvertent Algorithmic Cruelty. https://meyerweb.com/eric/thoughts/2014/12/24/inadvertent-algorithmic-cruelty/. Accessed: 2020-09-08.
[103] Scott M Monroe. 1982. Life events assessment: Current practices, emerging trends. *Clinical Psychology Review* 2, 4 (1982), 435–453.
[104] Elizabeth M Morgan, Chareen Snelson, and Patt Elison-Bowers. 2010. Image and video disclosure of substance use on social media websites. *Computers in Human Behavior* 26, 6 (2010), 1405–1411.
[105] Dhiraj Murthy. 2012. Towards a sociological understanding of social media: Theorizing Twitter. *Sociology* 46, 6 (2012), 1059–1073.
[106] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. 2010. Is it really about me?: message content in social awareness streams. In *Proc. CSCW*. ACM.
[107] James S Nairne, Sarah R Thompson, and Josefa NS Pandeirada. 2007. Adaptive memory: Survival processing enhances retention. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition 33, 2 (2007), 263.
[108] Michael D Newcomb, George J Huba, and Peter M Bentler. 1981. A multidimensional assessment of stressful life events among adolescents: Derivation and correlates. *Journal of health and social behavior* (1981), 400–415.
[109] Mark W Newman, Debra Lauterbach, Sean A Munson, Paul Resnick, and Margaret E Morris. 2011. It's not that i don't have problems, i'm just not putting them on facebook: challenges and opportunities in using online social networks for health. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 341–350.
[110] Hyun Jung Oh, Carolyn Lauckner, Jan Boehmer, Ryan Fewins-Bliss, and Kang Li. 2013. Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites. *Computers in Human Behavior* 29, 5 (2013), 2072–2080.
[111] Chester L Olson. 1979. Practical considerations in choosing a MANOVA test statistic: A rejoinder to Stevens. (1979).
[112] Julia Omarzu. 2000. A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review* 4, 2 (2000).
[113] Eileen YL Ong, Rebecca P Ang, Jim CM Ho, Joylynn CY Lim, Dion H Goh, Chei Sian Lee, and Alton YK Chua. 2011. Narcissism, extraversion and adolescents' self-presentation on Facebook. *Personality and individual differences* 50, 2 (2011), 180–185.
[114] Tuğçe Ozansoy Çadırcı and Ayşegül Sağkaya Güngör. 2019. Love my selfie: selfies in managing impressions on social networks. *Journal of Marketing Communications* 25, 3 (2019), 268–287.
[115] Xinru Page, Pamela Wisniewski, Bart P Knijnenburg, and Moses Namara. 2018. Social media's have-nots: an era of social disenfranchisement. *Internet Research* (2018).
[116] Harshvardhan J Pandit and Dave Lewis. 2018. Ease and ethics of user profiling in black mirror. In *Companion Proceedings of the The Web Conference 2018*.
[117] Galen Panger. 2014. Social comparison in social media: A look at Facebook and Twitter. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 2095–2100.
[118] Trishna Patel, Chris R Brewin, Jon Wheatley, Adrian Wells, Peter Fisher, and Samuel Myers. 2007. Intrusive images and memories in major depression. *Behaviour research and therapy* 45, 11 (2007), 2573–2580.
[119] Jessica Payne, Alexis M Chambers, and Elizabeth A Kensinger. 2012. Sleep promotes lasting changes in selective memory for emotional scenes. *Frontiers in integrative neuroscience* 6 (2012), 108.
[120] Sandra Petronio. 2002. *Boundaries of privacy: Dialectics of disclosure*. Suny Press.
[121] Pew. 2019. pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018.
[122] James Pierce. 2012. Undesigning technology: considering the negation of design by design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 957–966.
[123] Teri Quatman and Connie Swanson. 2002. Academic self-disclosure in adolescence. *Genetic, social, and general psychology monographs* 128, 1 (2002), 47–76.
[124] Leonard Reinecke and Sabine Trepte. 2014. Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. *Computers in Human Behavior* 30 (2014), 95–102.
[125] Michelle Richey, Aparna Gonibeed, and MN Ravishankar. 2018. The perils and promises of self-disclosure on social media. *Information Systems Frontiers* 20, 3 (2018), 425–437.
[126] Heidi R Riggio and Ronald E Riggio. 2002. Emotional expressiveness, extraversion, and neuroticism: A meta-analysis. *Journal of Nonverbal Behavior* 26, 4 (2002), 195–218.
[127] William Roberts and Janet Strayer. 1996. Empathy, emotional expressiveness, and prosocial behavior. *Child development* 67, 2 (1996), 449–470.
[128] Laura Robinson. 2007. The cyberself: the self-ing project goes online, symbolic interaction in the digital age. *New Media & Society* (2007).
[129] Koustuv Saha et al. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In *CHI Ext. Abstracts*. ACM.
[130] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *PACM Human-Computer Interaction* CSCW (2017).
[131] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *PACM HCI* CSCW (2019).
[132] Koustuv Saha, Manikanta D Reddy, Stephen Mattingly, Edward Moskal, Anusha Sirigiri, and Munmun De Choudhury. 2019. Libra: On linkedin based role ambiguity and its relationship with wellbeing and job performance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
[133] Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, and Munmun De Choudhury. 2019. A Social Media Study on the Effects of Psychiatric Medication Use. In *ICWSM*.
[134] Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In *ICWSM*.
[135] Ed Sandvik, Ed Diener, and Larry Seidlitz. 2009. Subjective well-being: The convergence and stability of self-report and non-self-report measures. In *Assessing well-being*. Springer, 119–138.
[136] Dietram A Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication* 57, 1 (2007), 9–20.
[137] Karen B Schmaling and Thomas L Patterson. 2019. The association of major life events with chronic fatigue. *Journal of psychosomatic research* (2019).
[138] Johann Schrammel, Christina Köffel, and Manfred Tscheligi. 2009. Personality traits, usage patterns and information disclosure in online communities. *People and Computers XXIII Celebrating People and Technology* (2009), 169–174.
[139] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*. Springer, 157–180.
[140] Judith A Scully, Henry Tosi, and Kevin Banning. 2000. Life event checklists: Revisiting the social readjustment rating scale after 30 years. *Educational and psychological measurement* 60, 6 (2000), 864–876.
[141] Pavica Sheldon. 2013. Examining gender differences in self-disclosure on Facebook versus face-to-face. *The Journal of Social Media in Society* 2, 1 (2013).
[142] Walter C Shipley. 2009. *Shipley-2: manual*. WPS.
[143] Christopher J Soto and Oliver P John. 2017. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology* 113, 1 (2017), 117.
[144] Charles D Spielberger, Fernando Gonzalez-Reigosa, Angel Martinez-Urrutia, Luiz FS Natalicio, and Diana S Natalicio. 2017. The state-trait anxiety inventory. *Revista Interamerican Journal of Psychology* (2017).
[145] Arati Srinivasan, Hong Guo, and Sarv Devaraj. 2017. Who cares about your big day? Impact of life events on dynamics of social networks. *Decision Sciences* 48, 6 (2017), 1062–1097.
[146] Anselm Strauss and Juliet Corbin. 1990. Open coding. *Basics of qualitative research: Grounded theory procedures and techniques* (1990).
[147] Diana I Tamir and Jason P Mitchell. 2012. Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences* 109, 21 (2012), 8038–8043.
[148] Samuel Hardman Taylor, Jevan Alexander Hutson, and Tyler Richard Alicea. 2017. Social consequences of Grindr use: Extending the Internet-enhanced self-disclosure hypothesis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6645–6657.
[149] Christopher Tennant. 2002. Life events, stress and depression: a review of recent findings. *Australian & New Zealand Journal of Psychiatry* 36, 2 (2002), 173–182.
[150] Winston Jin Song Teo and Chei Sian Lee. 2016. Sharing Brings Happiness?: Effects of Sharing in Social Media Among Adult Users. In *International Conference on Asian Digital Libraries*. Springer, 351–365.
[151] Robert P Tett, Douglas N Jackson, and Mitchell Rothstein. 1991. Personality measures as predictors of job performance: A meta-analytic review. *Personnel psychology* 44, 4 (1991), 703–742.
[152] Peggy A Thoits. 1983. Dimensions of life events that influence psychological distress: An evaluation and synthesis of the literature. In *Psychosocial stress*. Elsevier, 33–103.
[153] Edmund R Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology* 38, 2 (2007), 227–242.
[154] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response*.
[155] Zeynep Tufekci. 2008. Grooming, gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate? *Information, Communication & Society* 11, 4 (2008), 544–564.
[156] Jean M Twenge and Brian H Spitzberg. 2020. Declines in non-digital social interaction among Americans, 2003–2017. *Journal of Applied Social Psychology* (2020).
[157] Philip M Ullrich and Susan K Lutgendorf. 2002. Journaling about stressful events: Effects of cognitive processing and emotional expression. *Annals of Behavioral Medicine* 24, 3 (2002), 244–250.
[158] Niels van Berkel, Jorge Goncalves, Lauri Lovén, Denzil Ferreira, Simo Hosio, and Vassilis Kostakos. 2019. Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* 125 (2019), 118–128.
[159] Jessica Vitak. 2012. The Impact of Context Collapse and Privacy on Social Network Site Disclosures. *Journal of Broadcasting & Electronic Media* 56 (2012).
[160] Jessica Vitak and Jinyoung Kim. 2014. "You can't block people offline": examining how facebook's affordances shape the disclosure process. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*.
[161] Jessica Vitak, Cliff Lampe, Rebecca Gray, and Nicole B Ellison. 2012. " Why won't you be my Facebook friend?" strategies for managing context collapse in the workplace. In *Proceedings of the 2012 iConference*. 555–557.

[162] Yiran Wang, Melissa Niiya, Gloria Mark, Stephanie M Reich, and Mark Warschauer. 2015. Coming of Age (Digitally): An Ecological View of Social Media Use among College Students. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*.
[163] George J Warheit. 1979. Life events, coping, stress, and depressive symptomatology. *American Journal of Psychiatry* 136, 4B (1979), 502–507.
[164] David Watson and Lee Anna Clark. 1999. The PANAS-X: Manual for the positive and negative affect schedule-expanded form. (1999).
[165] Anita Whiting and David Williams. 2013. Why people use social media: a uses and gratifications approach. *Qualitative Market Research: An International Journal* (2013).
[166] Sally ME Wyatt. 2003. Non-users also matter: The construction of users and non-users of the Internet. *Now users matter: The co-construction of users and technology* (2003), 67–79.
[167] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2018. Detecting personal life events from twitter by multi-task lstm. In *Companion Proceedings of the The Web Conference 2018*. 21–22.

# A APPENDIX

## DISENTANGLING FACTORS OF REPORTING LIFE EVENTS ON DIFFERENT MODALITIES

Besides the convergence ($\mathbf{Model_1}$) and divergence models ($\mathbf{Model_2}$) as studied in Section 5.3, we also run a third kind of logistic regression models on the entire data of $\mathtt{D_T}$, such that:

- $\mathbf{Model_{3a}}$ uses all the described covariates as dependent variable and predicts if the event is disclosed on social media as the dependent variable, i.e., 1 if self-disclosed on social media, and 0 if not.
- $\mathbf{Model_{3b}}$ uses all the described covariates as dependent variable and predicts if the event is reported on survey as the dependent variable, i.e., 1 if reported on survey, and 0 if not.

Essentially, these models allow us to disentangle the effects of each of our covariates in explaining the direction of reporting, treating each of the modalities independent of each other. For instance, $\mathtt{Model_2}$ revealed that males show a negative correlation (Table 7) which could either be because males tend to disclose lesser on social media, or because Males report more on surveys compared to females. The two models $\mathbf{Model_{3a}}$ and $\mathbf{Model_{3b}}$ would help us to disentangle similar directions of the factors in each of the models.

Table A1 shows standardized coefficients and significance of the covariates in the above models. Looking at the significant variables, we find that an interesting pattern that $\mathbf{Model_{3a}}$ and $\mathbf{Model_{3b}}$ show coefficients with opposite signs. For example, age shows positive association with social media disclosures and a negative association with survey self-reports. Again, males are less likely to disclose events on social media, and, age has no effect on self-reports. We also find that healthy sleep quality has a strong negative association with social media disclosures, however no significant association with self-reports of life events.

Among event attributes, we find that valence of event bears a strong positive association with social media disclosures but no significant relationship with self-reports. In contrast, greater the significance of an event, less likely it is to be disclosed on social media, and more likely it is to be reported in self-reported survey. We construe similar explanation as in Section 5 holds here, significant events could be associated with emergency circumstances when the individual has lower propensity to post about the event. Similar associations are observed for recency, intimacy, and scope, with negative association with social media disclosure and positive association with self-reports. With respect to type of events, Work shows significant negative relationship with social media disclosure and positive relationship with self-reports — indicating that work related events are less likely to be posted on social media despite their occurrences.

Finally, we also find interesting directions for the baseline attributes, we find that social media related baseline attributes positively associate with social media disclosure but show no statistical significance in the relationship with survey based disclosure. For survey related baseline attributes, we find that number of survey records negatively associate with number of social media disclosures, and positively associate with survey event logging. Again, baseline self-reported significance shows a positive association with social media disclosure, indicating that individuals who tend to self-perceive greater significance of events are also more likely to disclose the event on social media. Taken together, the relationships observed in this analysis is not very different from what we observe in our results, providing more insight about what does the factors associated with online disclosures of life events.

**Table A1: Model$_{3*}$: Coefficients of linear regression of relevant covariates as independent variables and disclosing on social media as dependent variable in Model$_{3a}$ (1 for disclosure and 0 for no-disclosure), and self-reporting on survey as dependent variable in Model$_{3b}$ (1 for self-report and 0 for no-self-report), * $p$<0.05, ** $p$<0.01, *** $p$<0.001.**

| | Model$_{3a}$ | | Model$_{3b}$ | |
|---|---|---|---|---|
| **Demographic/Trait** | **Coeff.** | **p** | **Coeff.** | **p** |
| Age | 0.02 | * | −0.04 | *** |
| Gender: Male | −0.89 | * | 0.43 | *** |
| Born in US: Yes | −0.35 | | 0.22 | |
| Education: H. School | −0.02 | | 1.16 | *** |
| Education: College | −0.04 | | 1.19 | * |
| Education: Grad School | 0.22 | | 0.94 | ** |
| Education: Doctoral | 0.32 | | 0.72 | |
| Shipley: Abstraction | −0.07 | *** | 0.04 | * |
| Shipley: Vocabulary | −0.02 | | 0.05 | ** |
| Personality: Openness | 0.06 | | −0.34 | ** |
| Personality: Conscientiousness | −0.06 | | −0.07 | * |
| Personality: Extraversion | 0.22 | * | −0.03 | |
| Personality: Agreeableness | 0.92 | *** | 0.08 | |
| Personality: Neuroticism | 0.26 | * | −0.03 | |
| Positive Affect | −0.00 | | −0.03 | ** |
| Negative Affect | −0.00 | | 0.00 | |
| Stai: Anxiety | 0.02 | | −0.03 | * |
| PSQI: Healthy Sleep Quality | −0.10 | *** | 0.01 | |

| | Model$_{3a}$ | | Model$_{3b}$ | |
|---|---|---|---|---|
| **Event Attribute** | **Coeff.** | **p** | **Coeff.** | **p** |
| Valence: Positive | 0.39 | *** | −0.19 | |
| Significance | −1.26 | *** | 0.87 | *** |
| Recency | −1.88 | | 1.57 | *** |
| Anticipated | 0.19 | * | −0.16 | |
| Intimacy | −0.78 | *** | 0.36 | *** |
| Scope | −0.84 | *** | 0.44 | ** |
| Status: Ongoing | 4.71 | *** | −1.92 | *** |
| Type: Health | −0.25 | | 0.29 | |
| Type: Work | −1.54 | *** | 0.97 | *** |
| Type: School | −0.10 | | 0.26 | |
| Type: Local | −1.08 | * | −0.17 | |
| Type: Financial | −2.88 | *** | 1.35 | *** |
| **Baseline Attribute** | **Coeff.** | **p** | **Coeff.** | **p** |
| SM: Num. Posts | 1.28 | *** | -0.03 | |
| SM: Avg. Post Length | 8.62 | * | -1.44 | |
| SR: Num. Records | -1.07 | *** | 1.46 | *** |
| SR: Avg. Significance | 0.50 | *** | -0.17 | ** |

**Model$_{3a}$** : AIC = 920.3, Deg. Freedom= 34, LLk. = −425.15, $\chi^2$= 2493.88, Pseudo $R^2$ = 0.75, p < 0.001 ***
**Model$_{3b}$** : AIC = 2160.4, Deg. Freedom= 34, LLk. = −1045.22, $\chi^2$= 1468.90, Pseudo $R^2$ = 0.43, p < 0.001 ***