A Social Media Study on the Effects of Psychiatric Medication Use

Koustuv Saha[†], Benjamin Sugar[†], John Torous[‡], Bruno Abrahao[*], Emre Kıcıman[§], Munmun De Choudhury[†]

[†]Georgia Tech, [‡]Harvard Medical School, [*]NYU Shanghai, [§]Microsoft Research

[†]{koustuv.saha,bsugar,munmund}@gatech.edu, [‡]jtorous@bidmc.harvard.edu, [*]bd58@nyu.edu, [§]emrek@microsoft.com

Abstract

Understanding the effects of psychiatric medications during mental health treatment constitutes an active area of inquiry. While clinical trials help evaluate the effects of these medications, many trials suffer from a lack of generalizability to broader populations. We leverage social media data to examine psychopathological effects subject to self-reported usage of psychiatric medication. Using a list of common approved and regulated psychiatric drugs and a Twitter dataset of 300M posts from 30K individuals, we develop machine learning models to first assess effects relating to mood, cognition, depression, anxiety, psychosis, and suicidal ideation. Then, based on a stratified propensity score based causal analysis, we observe that use of specific drugs are associated with characteristic changes in an individual's psychopathology. We situate these observations in the psychiatry literature, with a deeper analysis of pre-treatment cues that predict treatment outcomes. Our work bears potential to inspire novel clinical investigations and to build tools for digital therapeutics.

Introduction

Psychiatric medications are key to treat many mental health conditions, including mood, psychotic, and anxiety disorders. 1 in 6 Americans take psychiatric medications and they account for 5 of the top 50 drugs sold in the U.S (*drugs.com*). These drugs[1] are designed to correct underlying neuro-pathological disease processes by restoring neural communication by modulating the brains chemical messengers and neurotransmitters (Barchas and Altemus 1999). These changes can be accompanied by debilitating neurological impairments and life-threatening effects as severe as suicidal ideation (Coupland et al. 2011) which reduce psychosocial functioning, and make social capital and vocational development less available to these individuals. Given the pervasiveness of their use, psychiatric medications can either alleviate or exacerbate mental illness burden on both personal and societal levels (Rosenblat et al. 2016).

One reason behind the mixed success of psychiatric medications stems from the fact that the mechanisms by which they modify the brain operation are poorly understood. In practice, their effects vary across individuals, and often do not achieve the intended result. Without any biological markers to match patients with the most appropriate medication, the selection of drug treatments is based primarily on trial-and-error (Cipriani et al. 2018; Trivedi et al. 2006). Unsurprisingly, frustration with treatment and side effects often causes treatment discontinuation (Bull et al. 2002).

Consequently, literature in precision psychiatry has emphasized the need to understand the psychiatric effects of these medications (Cipriani et al. 2009). Presently, most knowledge of drug reactions comes from clinical trials and reports of adverse events; e.g., the FDA's Adverse Event Reporting System (*open.fda.gov/data/faers*) clinical trial database. However, these trials can be biased, being conducted and funded by pharmaceutical companies, and are rarely replicated in large populations (Lexchin et al. 2003). In addition, these clinical trials suffer from limitations such as non-standardized study design, confounding factors, and restrictive eligibility criteria (Lexchin et al. 2003). For example, an analysis found that existing inclusion criteria for most trials would exclude 75% of individuals with major depressive disorders (Blanco et al. 2008). Even well-designed clinical trials can suffer from low statistical power, or limited observability of effects due to short monitoring and study periods, spanning just weeks or months

Contributions Our work seeks to address these gaps and complements existing methodologies for understanding the effects of psychiatric medications. We report a large-scale social media study of the effects of 49 FDA approved antidepressants across four major families (SSRIs, SNRIs, TCAs, and TeCAs) (descriptions in (Lopez-Munoz and Alamo 2009)). Our analysis is conducted using two years of Twitter data from two populations: 112M posts from 30K self-reported users of psychiatric medications and 707M posts from 300K users who did not. Adopting a patient-centered approach (Shippee et al. 2012), in this paper, we seek to study the effects of these drugs as reflected and self-reported in the naturalistic social media activities of individuals.

Accomplishing this goal involves meeting several technical challenges, importantly addressing causality, and our work offers robust and validated computational methods for the purpose. We first develop expert-validated machine learning models to assess psychopathological states

[1]This paper uses *medications* and *drugs* interchangeably, referring to U.S. FDA regulated psychiatric drugs only.

known to be affected by psychiatric medications, including mood, cognition, depression, anxiety, psychosis, and suicidal ideation, as given in the literature (Coupland et al. 2011). Using initial social media mentions of drug intake, we then identify individuals likely beginning treatment. Based on a stratified propensity score analysis (Olteanu et al. 2017), we compare post-treatment symptoms in treated individuals to large untreated control population. With an individual treatment effect analysis, we study the relationship between pre-treatment mental health signals and post-treatment response.

Findings Our results show that most drugs are linked to a post-treatment increase in negative affect and decrease in positive affect and cognition. We find varying effects both within and between the drug families on psychopathological symptoms (depression, anxiety, psychosis, and suicidal ideation). Clinically speaking, SSRIs are associated with worsening symptoms, whereas TCAs lead to improvements. Studying the individual-specific outcomes, our analyses help associate drug effectiveness with individuals' psycholinguistic attributes on social media.

Clinically, our findings reveal signals of the most common effects of the psychiatric medications over a large population, with the potential for improved characterization of their occurrence. Technologically, we show the potential of novel technologies in digital therapeutics, powered by large-scale social media analyses, to support digital therapeutics (Vieta 2015). These tools can improve the identification of adverse outcomes, as well as the behavioral and lifestyle changes in the heterogeneous outcomes of psychiatric drugs.

Privacy, Ethics, and Disclosure Given the sensitive nature of our work, despite working with public social media data, we are committed to securing the privacy of the individuals in our dataset. We use paraphrased examples of content and avoid personally identifiable information. Our findings were corroborated with our co-author who is a board-certified psychiatrist. *However, our work is not intended to replace clinical evaluation by a medical professional, and should not be used to compare or recommend medications.*

Background and Related Work

Psychiatric Drug Research and Prescriptions The mechanisms of action of many psychiatric drugs and the basis for specific therapeutic interventions, are not fully understood. Among other hypotheses, the monoanime hypothesis postulates that these drugs target the neurotransmitters serotonin, norepinephrine and dopamine, associated with feelings of well-being, alertness, and pleasure (Barchas and Altemus 1999). From the monoamine standpoint, medications are classified into families, based on their brain receptor affinities, which distinguish their mechanism of action.

Antidepressant research has grown tremendously, ever since Imipramine, and other Tricyclic Antidepressants (TCAs) were discovered and found to be effective (Gillman 2007). However, TCAs have a broad spectrum of neurotransmitter affinities, which may often lead to undesirable side effects, such as liver toxicity, excessive sleepiness, and sexual dysfunction (Frommer et al. 1987). Several other compounds have since been introduced whose development was guided by the idea that increasing the selectivity of the target of action to individual neurotransmitters would, in theory, limit the incidence of side effects while maintaining the effectiveness of TCAs (Lopez-Munoz and Alamo 2009). These include Tetracyclic Antidepressants (TeCA), Serotonin Norepinephrine Reuptake Inhibitors (SNRI), and Selective Serotonin Reuptake Inhibitors (SSRI).

Given these biochemical underpinnings, historically psychiatric care has adopted a "Disease-Centered Model" (Moncrieff and Cohen 2009), one that justifies prescribing medications on the assumption that they help correct the biological abnormalities related to psychiatric symptoms. However, this model neglects the psychoactive effects of the drugs. Consequently, a "Drug-Centered Model" has been advocated (Moncrieff and Cohen 2009), enabling patients to exercise more control over their pharmacotherapy, and moving treatment in a collaborative direction between clinicians and patients. Our work builds on this notion towards a "Patient-Centered Model" (Shippee et al. 2012), where psychiatrists could leverage complementary techniques (such as stratifying users on their naturalistic digital footprints) to prescribe medications.

Understanding Effects of Psychiatric Drugs The efficacy, safety, and approval of psychiatric drugs are typically established through clinical trials. In one such trial, the randomized controlled trial (RCT), participants are randomly assigned to a treatment or a control group, where the former receives a particular drug, and the latter receives a placebo (eg. a sugar pill with no drug content). Then, the effects of the treatment are measured as a difference in the two groups following the drug intake. A major weakness of these trials is that they are often conducted on individuals who may significantly differ from actual patients, and often, they are not externally validated to a larger and a more representative population (Hannan 2008). As an alternative, a study design that has gained interest is observational study (Hannan 2008). The advantage here is that they enable the researchers to conduct subset analyses that can help to precisely identify which patients benefit from each treatment. Similarly, we use large-scale longitudinal data and a causal approach to not only examine the effects of psychiatric drugs, but also to provide a framework that finds insights about their effectiveness across strata of populations.

Pharmacovigilance, Web, and Social Media

Pharmacovigilance is "the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem" (WHO 2002). Over the years, pharmacovigilance has become centered around data mining of clinical trial databases and patient-reported data. Recently, patient-generated activity online has also been used to understand pharmacological effects in large populations (Harpaz et al. 2017). White et al. (2016) found that web search logs improve detection of adverse effects by 19%, compared to an offline approach.

Social media studies of drug and substance use, including behavioral changes, adverse reactions, and recovery have

garnered significant attention in HCI (Chancellor et al. 2019, Kıcıman et al. 2018, Liu et al. 2017). Recent research has studied the abuse of prescription drugs, by leveraging drug forums (MacLean et al. 2015), Twitter (Sarker et al. 2015), and Reddit (Gaur et al. 2018). Social media has also facilitated the identification of adverse drug reactions at the population level using self-reports (Lardon et al. 2015) as well as the mentions of side effects of adverse drug reaction on Twitter (Nikfarjam et al. 2015).

Social media enables individuals to candidly share their personal and social experiences (Kıcıman et al. 2018, Olteanu et al. 2017, Saha et al. 2019b), thereby providing low-cost, large-scale, non-intrusive data to understand naturalistic patterns of mood, behavior, cognition, social milieu, and even mental and psychological states, both in real-time and longitudinally (Chancellor et al. 2016, Coppersmith et al. 2014, De Choudhury et al. 2013, Dos Reis and Culotta 2015, Saha et al. 2019a, Yoo and De Choudhury 2019). In characterizing drug use, being able to quantify these psychopathological attributes is extremely powerful.

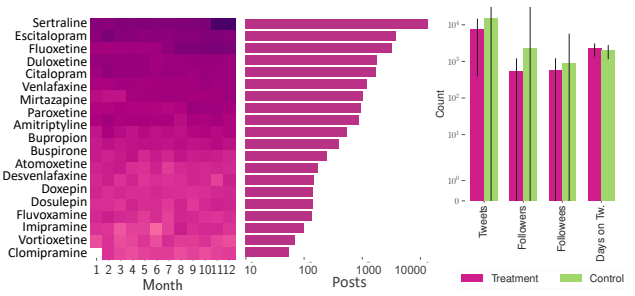
Nevertheless, we observe a gap that digital pharmacovigilance studies, particularly those using social media, have largely targeted the named adverse effects of drugs (e.g., "headache", "palpitations", "nausea"), and have not measured broader forms of symptomatic changes longitudinally. To fill this gap, our work draws on theoretically grounded methodologies, including lexicon-based and machine learning approaches, to measure the symptomatic outcomes of psychiatric drug use longitudinally, including mood, cognition, depression, anxiety, psychosis, and suicidal ideation.

Data

This work leverages Twitter timeline data of individuals who self-report their use of psychiatric medications. The data collection involve: 1) curating a list of psychiatric medications; 2) using this list to collect Twitter posts that mentioned these medications; 3) identifying and filtering for only those posts where users self-reported about personal medication intake (using a personal medication intake classifier, and 4) collecting the timeline datasets of these individuals who self-reported psychiatric medication intake, and additionally doing that for another set of users who did not self-report psychiatric medication intake. We explain these steps here:

Psychiatric Medication List We scope our work to a list of FDA approved antidepressants and antidepressant augmentation drugs. We crawl a hand-curated set of Wikipedia pages of these drugs, to collect brand names, generic names, and drug family information to obtain a list of 297 brand names mapped to 49 generic names, grouped into four major families: SNRI, SSRI, TCA, TeCA. Our clinician co-author established the validity and relevance of this final list.

Twitter Data of Psychiatric Medication Usage We query the Twitter API for public English posts mentioning these drugs (brand or generic name) between January 01, 2015 and December 31, 2016 to obtain 601,134 posts by 230,573 unique users. A two year period balances concerns about being long enough to avoid confounds by idiosyncratic events



(a) (b)

Figure 1: (a) Monthly distribution and the number of posts in logarithmic scale for the top 20 medications (darker colors correspond to greaeter density); (b) Mean distribution of User Attributes in *Treatment* and *Control* datasets.

I'm taking my first dose of X tonight.
I was depressed & psychiatrist gave me X, slept for two days!
First day on X. Dose 1 taken, and I already feel weird from it.
Just took X for the first time. Let's see how it goes
I got brain zaps if I took X_1 even an hour late. Changed to X_2 now!
My no-med experiment went horribly awry, so I'm starting X today

Table 1: Example paraphrased self-reports of psychiatric medication usage. Drug names are masked.

and seasonal changes, but short enough to avoid major changes in social media use and drug prescription policies. This also enables us to collect sufficient pre- and post- medication usage timeline data for our ensuing analyses.

Personal Medication Intake Classifier Since mentioning a medication in a tweet does not necessarily indicate its usage, we filter out those posts that were first-person reports of using these medications. For this purpose, we employ a machine learning classifier built in a recent work (Klein et al. 2017). This classifier distinguishes Twitter posts into the binary classes (yes or no) if there is a self-report about personal medication intake. We replicate this model and train it on an expert-annotated dataset of 7,154 Twitter posts (dataset published in Klein et al. 2017). The classifier uses an SVM model with linear kernel and shows a mean k-fold (k=5) cross-validation accuracy and F1-score of 0.82 each.

We use this classifier to label the 601,134 medication-mention posts to find that 93,275 of these posts indicate medication self-intake (example posts in Table 1). Figure 1a shows the monthly and overall distribution of the top 20 drugs in our dataset. We find that SSRIs (eg. Sertraline, Escitalopram, Fluoxetine) rank highest in the distribution. This aligns with external surveys on the most prescribed psychiatric drugs in that time which found that the top 5 antidepressants captured over 70% of the prescription volumes (Scripts 2018; Gohol 2018).

Compiling Treatment and Control Datasets The above 93,275 medication usage posts were posted by 52,567 unique users from whom we then collect Twitter metadata such as the number of tweets, followers, followees, and account creation date. To limit our analyses to typical Twitter users, we remove users (e.g., celebrities or typically inactive

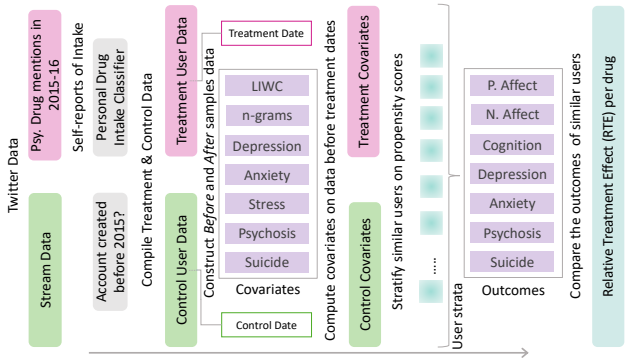


Figure 2: Schematic diagram of propensity score analysis.

users) with more than 5000 followers or followees or posted outside the range of 200 to 30,000 tweets—a choice motivated from prior work (Pavalanathan and Eisenstein 2016). For the remaining 34,518, we collect the timeline data between January 01, 2014 and February 15, 2018, to obtain a total of 112,025,496 posts. Finally, we limit our dataset to those users who posted both before and after their first self-reported use of medication and did not self-report the use before 2015. The resultant timeline dataset of 23,191 users is referred to hereon as the *Treatment* dataset.

Additionally, we build a *Control* dataset of users who did not self-report using psychiatric medication. We obtain 495,419 usernames via the Twitter streaming API and prune this list (as above) and remove accounts that did not exist pre-2015. We collect the timelines of the remaining 283,374 users, for a total 707,475,862 posts. Figure 1b shows the mean distribution of Twitter attributes in our two datasets.

Methods

Study Design and Rationale Recall that our research objective is to examine the effects of psychiatric medications in terms of the changes in mental health symptoms. Effectively answering this question necessitates the use of causal methods to reduce biases associated with the observed effects following the reported medication usage. The effects of drugs are most often measured through Randomized Controlled Trials (RCTs) in clinical settings (Cipriani et al. 2018; Szegedi et al. 2005). Due to the limitations of this approach, noted in the "Background and Related Work" section, and because of the potential advantages of a "Patient Centered Model" that focuses on using the naturalistic self-reports of individuals regarding their psychiatric medication use, this work adopts an observational study design. We do acknowledge that observational studies are weaker than RCTs in making conclusive causal claims like ones needed to accomplish the goals of this paper, but they provide complementary advantages over RCTs in many aspects (Hannan 2008). Literature in statistics also provides support for these methods and similar frameworks have been leveraged in previous quantitative social media studies (De Choudhury et al. 2016, Kıcıman et al. 2018, Olteanu et al. 2017, Saha et al. 2018).

Specifically, we adopt a causal inference framework based on matching, which simulates an RCT setting by controlling for as many covariates as possible (Imbens and Rubin 2015). This approach is built on the potential outcomes framework, which examines whether an outcome is caused by a treatment T, by comparing two potential outcomes: 1) $Y_i(T = 1)$ when exposed to T, and 2) $Y_i(T = 0)$ if there was no T. However, it is impossible to obtain both of these outcomes for the same individual. To overcome this challenge of missing data, this framework estimates the missing counterfactual outcome for an individual based on the outcomes of other similar (matched) individuals (in terms of their covariate distribution). In particular, we employ stratified propensity score analysis (Olteanu et al. 2017) to match and then to examine the symptomatic outcomes in the *Treatment* and *Control* individuals by measuring the relative treatment effect of the drugs (see Figure 2 for an overview).

Constructing *Before* and *After* Samples

As our setting concerns measuring the changes *post* reported usage of the medications, we divide our datasets into *Before* and *After* samples around their dates of treatment. For every *Treatment* user, we assign the date of their first medication-intake post as their treatment date. We assign each individual in the *Control* dataset a placebo date, matching the non-parametric distribution of treatment dates of the *Treatment* dataset, to mitigate the effects of any temporal confounds. For this, we ensure that the treatment and placebo dates follow similar distribution by non-parametrically simulating placebo dates from the pool of treatment dates. We measure the similarity in their distribution using Kolmogorov–Smirnov test to obtain an extremely low statistic of 0.06, indicating similarity in the probability distribution of treatment and placebo dates (Figure 3b). We then divided our *Treatment* and *Control* datasets into *Before* and *After* samples based on the treatment and placebo dates.

Defining and Measuring Symptomatic Outcomes

Drawing on the psychiatry and psychology literature (Pennebaker et al. 2003, Rosenblat et al. 2016), next, we measure mental health symptomatic outcomes, subject to the reported usage of the medications in the above-constructed user samples, based on the changes in mood, cognition, depression, anxiety, stress, psychosis, and suicidal ideation. We use the following approaches:

Affect and Cognition To measure the affective and cognitive outcomes, similar to prior work (Ernala et al. 2017, Saha et al. 2018), we quantify psycholinguistic shifts in affect and cognition. In particular, we use the changes in the normalized occurrences of words in these categories per the well-validated Linguistic Inquiry and Word Count (LIWC) lexicon (Tausczik and Pennebaker 2010). These categories include *positive* and *negative affect* for affect, and *cognition mechanics*, *causation*, *certainty*, *inhibition*, *discrepancies*, *negation*, and *tentativeness* for cognition.

Depression, Anxiety, Stress, Psychosis, Suicidal Ideation We quantitatively estimate these measures from social media by building several supervised learning based classifiers

of mental health attributes. Our approach is inspired by recent work where mental health attributes have been inferred in unlabeled data by transferring a classifier trained on a different labeled dataset (Saha and De Choudhury 2017). To train such classifiers for use in our work, we identify several Reddit communities that are most closely associated with these measures. That is, the positive examples in our training data comprise posts shared on *r/depression* for depression, *r/anxiety* for anxiety, *r/stress* for stress, *r/psychosis* for psychosis, and *r/SuicideWatch* for suicidal ideation. On the other hand, negative examples are extracted from the collated sample of 20M Reddit posts gathered from 20 subreddits that appear on the landing page of Reddit during the same period of our Twitter data sample, such as *r/AskReddit*,*r/aww*, *r/movies*, and others.

These classifiers are SVM models with linear kernels and use 5000 n-grams (n=1,2,3) as features. We use a *balanced* number of examples for the two classes in training, and we tune the parameters of the classifiers using k-fold (k=5) cross-validation (Chandrasekharan et al. 2018). Table 2 summarizes the size of the datasets and the accuracy metrics. Figure 3a shows the ROC curves of these classifiers. These classifiers show a mean cross-validation accuracy ranging between 0.79 and 0.88 and mean test accuracy ranging between 0.81 and 0.91. Table 3 reports the top 10 features in each of the classifiers. Several top n-gram features such as *depression*, *stress*, *hope*, *help*, and *feel*, are contextually related to mental health.

Establishing Model Validity. Since our next goal is to employ these classifiers, trained on Reddit data, to automatically infer the symptomatic outcomes in the Twitter user samples—a platform with distinct norms and posting style, we present a series of evaluation tests to demonstrate the validity of the transfer approach and the transferred classifiers. 1) First, motivated from prior work (Saha et al. 2017a), we conduct a linguistic equivalence test between the Reddit training dataset, and the Twitter unseen dataset based on a word-vector similarity approach. Using word-vectors (pre-trained on Google News dataset of over 100 billion tokens), we find the vector similarity of the top 500 n-grams in the Reddit and Twitter corpuses to be 0.95. This shows high content similarity across the two platforms, in turn justifying the transfer approach. 2) 2) Second, we find that the top features of these classifiers align with that of similar mental health classifiers built on Twitter to identify depression (De Choudhury et al. 2013), anxiety (Dutta et al. 2018), stress (Lin et al. 2014), psychosis (Birnbaum et al. 2017), and suicidal ideation (Burnap et al. 2015). This indicates the construct validity of the transferred classifiers. 3) Third, we demonstrate convergence and divergence validity and present a qualitative validation of the outputs of these classifiers. Two researchers manually inspected 170 randomly selected Twitter posts on mental health symptoms, spanning both user samples. Using the methodology outlined in Bagroy et al. (2017) that draws up the DSM-5 clinical framework, they rated each Twitter post on a binary Likert scale (high/low) to assess levels of expressed depression, anxiety, stress, psychosis, or suicidal ideation. We find high (87%) agreement between the manual ratings and the classifiers' respective labels. This aligns with prior work where similar agreements have been reached between classifier outcomes and annotations of mental health experts (a Fleiss' κ=0.84 was reported in Bagroy et al. (2017)).

Matching For Causal Inference

Matching Covariates When conditioned on high-dimensional covariate data, matching is known to significantly minimize bias compared to naive correlational analyses (Imbens and Rubin 2015). Our approach controls for a variety of covariates so that the compared *Control* and *Treatment* groups show similar pre-treatment online behavior. The 1st set of covariates includes users' *social attributes* (count of tweets, followers, followees, duration on the platform and frequency of posting). The 2nd set corresponds to the distribution of word usage in the Twitter timelines, where for every user, we build a vector model on the top 2,000 unigrams. The 3rd set consists of normalized use of psycholinguistic attributes in the posts, i.e, distribution across 50 categories in the LIWC lexicon (Tausczik and Pennebaker 2010), across *affective*, *cognitive*, *lexical*, *stylistic*, and *social* attributes.

Finally, to minimize the confounding effects of an individual's mental health conditions prior to treatment, in the 4th set we control for the users' mean aggregated usage of posts indicative of *depression*, *anxiety*, *stress*, *psychosis*, and *suicidal ideation*, assessed using the classifiers described above. Note that there is typically a significant time-lag between the onset of mental illness and the first treatment people receive (Hasin et al. 2005; Oliver et al. 2018). Therefore, matching on these pre-treatment symptoms should capture and account for the individual's already existing mental health condition. That is, our matched comparisons should on average be comparing people with a given mental illness who receive treatment to their counterparts who have the same symptoms but did not receive treatment.

Propensity Score Analysis We use matching to find pairs (generalizable to groups) of *Treatment* and *Control* users whose covariates are statistically very similar to one another, but where one was *treated*, and the other was not. The propensity score model matches users based on their *likelihood* of receiving the treatment, or the propensity scores. Our stratified matching approach groups individuals with similar propensity scores into strata (Kıcıman et al. 2018). Every stratum, therefore, consist of individuals with similar covariates. This helps us to isolate and estimate the effects of the treatment within each stratum.

To compute the propensity scores, we build a logistic regression model that predicts a user's treatment status based on their covariates. Next, we discard the outliers in the propensity scores (outside the range of 2 standard deviations from the mean), and segregate the remaining distribution into 100 strata of equal width. To further ensure that our causal analysis per stratum remains restricted to a sufficient number of similar users, we remove those strata with very few *Treatment* or *Control* users, as is common practice in causal inference research (De Choudhury et al. 2016). With a threshold of at least 50 users per group in a stratum, this

Precision		Recall		Accuracy	
CV	Test	CV	Test	CV	Test
Depression (40,000; 555,955)					
.88	.86	.88	.82	.88	.82
Anxiety (40,000; 238,689)					
.82	.91	.82	.90	.82	.91
Stress (5,000; 5,969)					
.79	.92	.79	.91	.79	.92
Psychosis (5,000; 3,439)					
.87	.85	.87	.81	.87	.81
Suicidal Idn. (40,000; 276,769)					
.78	.91	.78	.91	.78	.91

Table 2: Mental health classifiers (training:test data size), cross-validation and test accuracies.

Depression		Anxiety		Stress		Psychosis		Suicidal Idn.	
Feature	Score	Feature	Score	Feature	Score	Feature	Score	Feature	Score
concerns	.6	*forgetting*	.6	*stress*	.4	*psychosis*	.5	*help*	.4
it looks like	.5	*it looks*	.6	*help*	.4	*song*	.4	*friends*	.4
here are	.5	*does it*	.6	*try*	.4	*psychotic*	.4	*anymore*	.4
forgetting	.4	*looks like*	.6	*work*	.3	*hope*	.3	*never*	.4
know	.4	*concerns*	.6	*feel*	.3	*experience*	.3	*family*	.4
all really	.4	*posting*	.5	*things*	.2	*help*	.3	*suicide*	.4
depression	.4	*anxiety*	.4	*you can*	.3	*schizophenia*	.3	*people*	.4
have spaces	.3	*around*	.4	*life*	.2	*symptoms*	.3	*end*	.4
suicidal	.3	*feel*	14.5	*take*	.2	*medication*	.2	*think*	.3
feeling	.2	*attack*	.3	*need to*	.2	*weed*	.2	*around*	.3

Table 3: Top 10 Features in the mental health outcome classifiers.

approach gave 63 strata that consisted of 23,163 *Treatment* and 122,941 *Control* individuals (Figure 3c).

Quality of Matching To ensure that we matched statistically comparable *Treatment* and *Control* users, we evaluate the balance of their covariates. We compute the standardized mean difference (SMD) across covariates in the *Treatment* and *Control* groups in each of the 63 valid strata. SMD calculates the difference in the mean covariate values between the two groups as a fraction of the pooled standard deviation of the two groups. Two groups are considered to be balanced if all the covariates reveal SMD lower than 0.2 (Kıcıman et al. 2018), a condition which all our covariates satisfied. We also find a significant drop in the mean SMD from 0.029 (max=0.31) in the unmatched datasets to 0.009 (max=0.05) in the matched datasets (Figure 3d).

Characterizing the Propensity Strata of Users To understand how the subpopulations across the several strata vary, we characterize their psycholinguistic attributes. Figure 4 plots the usage of affective and cognitive words across all the strata. The propensity score model distributed these users in such a way that the users with a greater tendency to use affective and cognitive words mostly occur in the lower and middle strata, whereas those with a lower tendency to use these words predominantly occur in the higher strata.

Measuring Changes in the Outcomes. To quantify the effects of self-reported psychiatric medication use, we compute the change in the symptomatic outcomes, weighted on the number of *Treatment* users in each stratum. For this, we first determine the Relative Treatment Effect (*RTE*) of the drugs per outcome measure in every stratum, as a ratio of the likelihood of an outcome measure in the *Treatment* group to that in the *Control* group (Kıcıman et al. 2018). Next, using a weighted average across the strata, we obtain the mean *RTE* of the medications per outcome measure. We compute the mean *RTE* for all the drugs and aggregate that for the drug families. An outcome *RTE* greater than 1 suggests that the outcome *increased* in the *Treatment* users, whereas an *RTE* lower than 1 suggests that it decreased in the *Treatment* users, following the reported use of psychiatric medication.

Exploring Individual-Specific Effects

We finally aim to study how the drugs affect individuals who vary in their pre-existing psychological state. So once we calculated the treatment effect of the drugs, we explore its

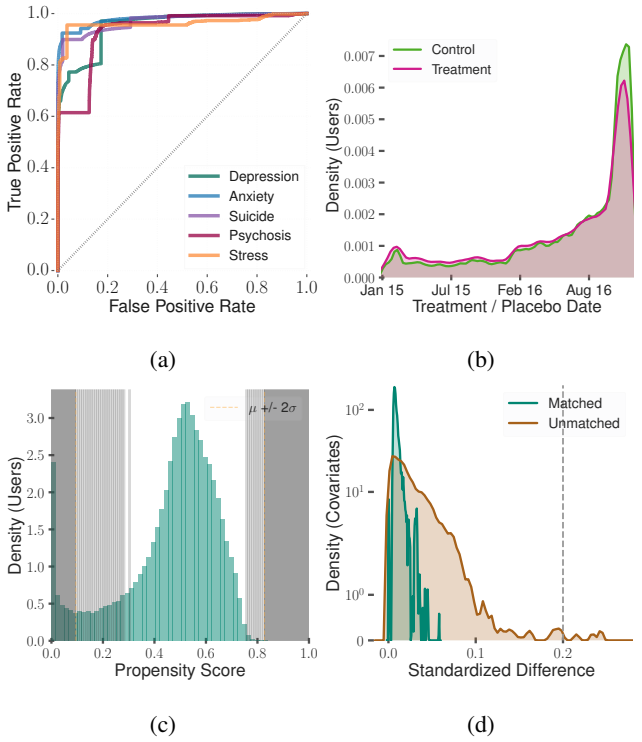
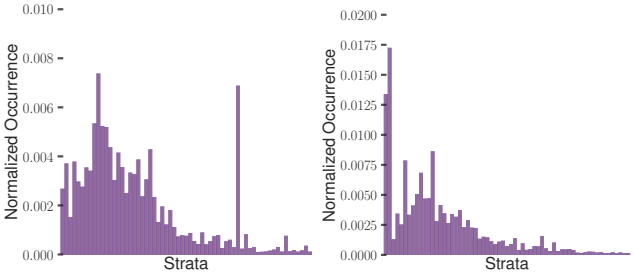


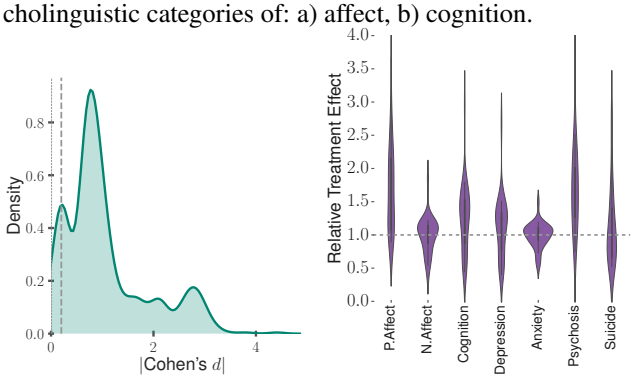
Figure 3: (a) ROC curves of the classifiers that measure symptomatic outcome, (b) *Treatment* dates distribution, (c) Propensity score distribution (shaded region represents those dropped in our analysis), (d) Quality of matching

relationship with the individuals' psycholinguistic attributes (as obtained by LIWC). For this, in every stratum, we first build separate linear regression models for all the outcomes of *Control* users with their covariates as predictors. Using these models we predict the counterfactual outcomes of the *Treatment* users in the strata – that is, the outcome for each treated user if they had not taken the drug. Next, for every user, we obtain the ratio of the predicted and actual value of the outcome. This essentially quantifies how much a *Treatment* user is individually effected by treatment, and is referred to as the Individual Treatment Effect (*ITE*) in individualized and precision medicine literature (Lamont et al. 2018). Finally, we measure the association between pre-treatment psycholinguistic attributes and the *ITE* values per



(a) (b)

Figure 4: Distribution of words by users across strata by psycholinguistic categories of: a) affect, b) cognition.



(a) (b)

Figure 5: (a) Distribution of effect size magnitude in the outcome change between *Treatment* and *Control* users; (b) Distribution of RTE across all the *Treatment* users.

drug, by fitting a linear regression model. This characterizes the directionality and the effect of a drug on an individual based on their pre-existing psycholinguistic attributes.

Results

Observations about Symptomatic Outcomes

Our first set of results investigates if self-reported psychiatric drug use had a statistically significant effect on the *Treatment* users. For this, we measure the effect size (Cohen's d) in the outcome changes between the *Treatment* and *Control* users, per drug, per outcome, and per valid strata. We find that the magnitude of Cohen's d averages at 0.75 (see Figure 5a). A cohen's d magnitude lower than 0.2 suggests small differences between two distributions. We find that 91% of our values fall outside this range, suggesting the *Treatment* significantly differed from the *Control* group. An independent sample t-test further reveals statistical significance in these differences ($t\epsilon$[-9.87,10.96]; p¡0.001), confirming that after the self-reported use of medications, the *Treatment* users showed significant changes in outcomes.

We then compute the Relative Treatment Effect (*RTE*) of the psychiatric medications. Figure 5b shows the distribution of *RTE* across the symptomatic outcomes for the matched *Treatment* and *Control* users. We find that the *RTE* across the outcomes averages at 1.28 (stdev=0.61). We dig deeper into the effects per drug. Figure 6 presents the *RTE* of the 20 most popular generic drugs and the 4 drug families. We

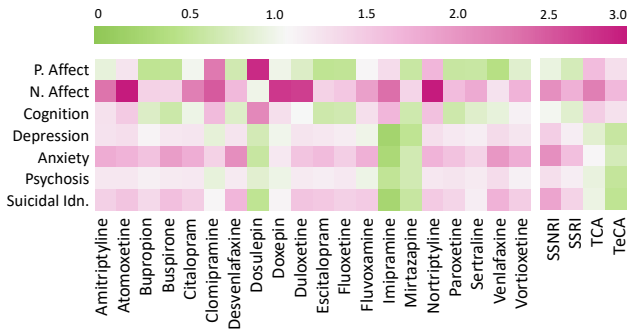
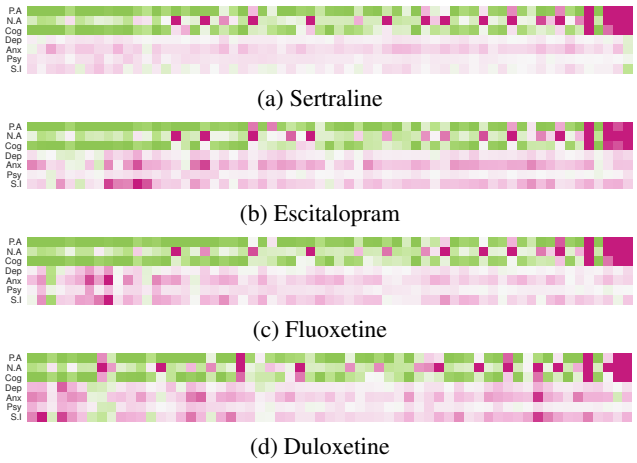


Figure 6: Relative Treatment Effect on the outcomes per 20 most popular drugs (left), and drug families (right).



(a) Sertraline

(b) Escitalopram

(c) Fluoxetine

(d) Duloxetine

Figure 7: *RTE* per propensity stratum for the top four drugs (For colorbar, refer to the one in Figure 6).

observe many interesting patterns here, such as most medications lead to similar directionality of effects on all the outcomes, e.g., all of the outcomes, depression, anxiety, psychosis, and suicidal ideation increase for the *Treatment* users in the *After* period of reported medication use. The similarity in effects across outcomes could be attributed to the comorbidity of the symptomatic outcomes and the clinical presentation of many moods and psychotic disorders (Rosenblat et al. 2016). We also observe that those drugs with similar pharmacological composition, such as Escitalopram and Citalopram, and Desvenlafaxine and Venlafaxine show similar trends in the symptomatic outcomes.

Table 4 summarizes the proportion of *Treatment* users who showed an increased outcome per drug family. For all these outcomes other than *positive affect* and *cognition* (in which case it is the opposite), an increase in the outcome measure also translates to *worsened observable mental health condition* of the individuals, whereas a decrease suggests an *improvement in their mental health condition*, as gleaned from Twitter. To study the strata-wise variation for each of these outcomes, we present Figure 7, which shows the *RTE* per stratum for the four most popular medications.

Effects on Affect and Cognition Figure 6 and Table 4 together indicate that the top medications and families are associated with an increase in the likelihood of negative affect.

Family	Users	P.A	N.A	Cog.	Dep.	Anx.	Psy.	S.I
SNRI	2535	21	57	33	81	93	76	83
SSRI	16388	19	59	30	78	98	79	94
TCA	2535	47	.52	51	35	62	33	36
TeCA	763	13	55	25	17	24	23	18

Table 4: Outcome measures per drug family, showing the percentage of users in strata showing *RTE* greater than 1.

However, that the likelihood of positive affect and cognition also decrease for most of these medications, aligns with literature about the inverse relationship observed in the occurrence of these attributes and mental health symptoms (Pennebaker, Mehl, and Niederhoffer 2003). Among the drug families, we find that the TCAs show the greatest improvement in these measures, with about half of their users showing increased positive affect and cognition.

Next, Figure 7 shows that these outcome measures decrease mostly in the lower-valued strata and increase in the higher valued ones (Figure 7). Note that these measures are not mutually exclusive. That is, an individual can see both increasing positive affect and increasing negative affect if they are using more affective words overall. The higher strata included users who typically showed lower affect and cognition than the rest (see Figure 4). Together, our findings suggest that the self-reported use of these medications is associated with ineffective (or worsening) effects on individuals with lower affective expressiveness and cognitive processing. Interestingly, these symptoms are also comorbid with mood disorders (Rosenblat et al. 2016), and the observed ineffectiveness of the drugs is likely influenced by the severity of their mental illness. However, to disentangle that requires further investigation, beyond the scope of our work.

Effects on Depression, Anxiety, Psychosis, and Suicidal Ideation For these second set of outcomes, we observe varied changes across medications. We observe that reported use of most of the medications are associated with worsening of these outcomes. These also include the most popular medications such as Sertraline, Escitalopram, and Fluoxetine. All of these are classified as SSRIs—the family which shows the most worsening in these outcomes among the drug families. In fact, our dataset reveals that within SSRIs, over 90% of the users were in strata that showed increased anxiety and suicidal ideation. On the other hand, we find improving symptoms in TCAs such as Dosulepin, Imipramine, and Clomipramine. From the perspective of drug families, the TCAs and the TeCAs show the greatest improving effects, with the majority of their users belonging to strata with decreased effects in the outcome measures.

Although most medications show similar effects at an aggregated level, we find differences in their strata-wise effects distributions (Figure 7). For example, in case of Duloxetine, we find minimal effects in the middle region, the one that showed high cognition (Figure 4). In contrast, Fluoxetine showed improving effects in a few lower valued strata. This observation—that the strata-wise effects can be different, inspired our next set of post-hoc analyses, wherein we examine individual-specific effects and drug-specific changes associated with the reported use of the medications.

Attribute	Coefficient	Attribute	Coefficient
Sertraline		Fluoxetine	
Past Tense	0.52	*Cognitive Mech.*	0.35
Tentativeness	0.35	*Present Tense*	0.34
1st P. Singular	-0.18	*Relative*	0.31
Aux. Verbs	-0.23	*Percept*	0.30
Cognitive Mech.	-0.25	*Conjunction*	-0.10
Escitalopram		Duloxetine	
Article	0.22	*Cognitive Mech.*	0.46
1st P. Singular	0.10	*Relative*	0.44
Social	-0.07	*1st P. Singular*	0.41
Bio	-0.13	*Social*	-0.20
2nd Person	-0.18	*Work*	-0.26

Table 5: Individual Treatment Effects: Relationship between pre-treatment attributes and improvement coefficient (Positive indicates *improvement*, Negative indicates *worsening*).

Understanding Individual-Treatment Effects

To understand how pre-treatment psycholinguistic signals correlate with post-treatment response to the drugs, we examine the effects at the individual level. For every *Treatment* user, we obtained their Individual Treatment Effect (*ITE*) values for all outcomes. Next, we fit several linear regression models per psychiatric medication to obtain the relationship between the *ITE*s and the psycholinguistic (LIWC) attributes of the users who reported using the medication. To simplify interpretability, corresponding to every psycholinguistic attribute, we averaged the coefficients of outcomes (preserving their directionality of improvement). For the four most popular drugs, Table 5 reports the coefficients of five psycholinguistic attributes with the greatest magnitudes in improvement or worsening. We summarize a few distinct patterns below, noted by our clinician coauthor to be most salient, based on the clinical literature and experience:

For Sertraline, the use of first person singular and auxiliary verb shows negative coefficients, indicating that this drug might not be effective in those with greater preoccupation and self-attentional focus—the known characteristics of these two attribute usage, typically prevalent in depressed individuals (De Choudhury et al. 2013). In contrast, Escitalopram and Duloxetine shows better efficacy in those individuals who have greater pre-occupation and lower social integration. Similarly, Fluoxetine and Duloxetine shows better efficacy in those individuals with greater usage of cognitive words—typically those who show lower cognitive impairment, but Sertraline shows the opposite effect in them.

Discussion

Our work presents two significant contributions: 1) By detecting the effects of drug use and that these changes are sensitive to drug families, we show a proof of concept that social media is useful as an effective sensor to scalably detect behavioral changes in individuals who initiate treatment via (self-reported) use of psychiatric medication; and 2) our empirical findings include the discovery that people's online behaviors change in some unexpected ways following drug intake, and these may differ from the named side-effects of these drugs. We discuss the significance and implications of these contributions in the remainder of this section.

Contextualizing the Findings in Psychiatry

As highlighted earlier, there are complexities in determining the effects of psychiatric medications in individuals; but at the same time, there are discrepancies in the claims made by clinical studies. For example, Geddes et al. found no major differences in the efficacy of SSRIs and TCAs, whereas other studies found one kind to perform better than others (Cipriani et al. 2018). Other studies found placebos or non-pharmacological care to have outperformed certain antidepressants (Szegedi et al. 2005). These conflicting findings in the literature prevent us from drawing conclusive claims about the validity of our findings.

From the perspective of clinical literature, our results offer varied interpretations. Figure 6 indicates a small impact of antidepressants on cognitive symptoms—an observation consistent with clinical experience and studies (Rosenblat et al., 2016). It is more difficult to explain the variable impact of the drugs on depressive symptoms. For instance, in our post-hoc analysis, Sertraline showed poor effects for individuals exhibiting attributes of depression, despite clinical evidence suggesting the opposite. On the other hand, Duloxetine was associated with positive symptomatic outcomes, as also found in clinical studies (Cipriani et al. 2018). Nevertheless, that these antidepressants have varying effects on individuals across strata finds support in clinical trials which report varying efficacy of antidepressants on different cohorts (Coupland et al. 2011)

Notwithstanding these varied findings, our work highlights the potential of older antidepressants. While TCAs (Imipramine, Clomipramine) are not often prescribed today because of serious toxicity issues that may be fatal in overdose (Kerr, McGuffie, and Wilkie 2001), our results demonstrate their effectiveness with the most favorable responses reported, compared to the other classes of anti-depressants.

Clinical Implications

Patient-Centered Approach to Pharmacological Care Our findings show that social media can provide valuable complementary insights into the effects of psychiatric drugs. This can complement clinical trials, allowing observations in larger populations and over longer time spans. Further, in psychiatry, medications are still prescribed by trial-and-error, or based on side effect profiles of these medications (Trivedi et al. 2006). Our analysis of individual treatment effects shows that the pre-treatment signals of mental health states appear to be linked to or predictive of individual drug success, raising the possibility of using such signals for **precision psychiatry** (Vieta 2015). While we use social media to demonstrate that this relationship exists, other sources of mental health signals may be used to complement our analyses, that are reliable and more broadly available.

Drug Repurposing Our results offer a novel opportunity to advance **drug repurposing**. Presently the pipeline for new pharmacological agents for mental illnesses is sparse (Dubovsky 2018), apart from ongoing research on ketamine and other potential new antidepressants (Dubovsky 2018). Drug repurposing— finding new clinical applications for currently approved medications, offering the potential of low cost and quicker to market treatments (Corsello et al. 2017). So far drug repurposing efforts in mental illnesses like depression have focused on biological targets (Powell et al. 2017). Although these approaches have been successful in identifying plausible repositioning candidates, a key challenge is providing direct evidence of candidate efficacy in people, rather than relying on surrogate biomarkers or indirect evidence. This is the first research to explore how social media may serve to identify novel targets as well. Our methods highlight how large quantities of real-time data can offer low cost and high volume assessments of people's own reports and perceptions related to antidepressants' use.

Technological Implications

Technologies for Regulatory Bodies Our results offer an important tool in generating "real-world evidence" for incorporation into technologies that can be used by regulatory bodies like the FDA. The FDA seeks to advance its approach to regulate and rely more on real-world evidence in addition to pre-market clinical studies data. As the FDA currently writes its novel digital health software program certification plan, where medical software such as smartphone applications will receive FDA approval without extensive clinical research—a key component is stated to be "monitoring real-world performance", though it is to be noted that they are still "considering how to best work to collect and interpret information about the product's safety and effectiveness" (*fda.gov*). This paper offers a novel technological approach that may meet the evolving needs of the FDA, by being able to identify the uses and effects of various medications as self-reported by people on social media.

Technologies for Drug Safety Surveillance From a public health perspective, our methods offer the potential to build technologies that surface early warning signs of adverse effects related to psychiatric drug use. The FDA's current Sentinel Initiative which aims to apply big data methods to medical claims data from over 5.5 billion patient encounters in an effort to flag previously unrecognized drug safety issues and to tackle issues of under-reporting of drug effects, has still not superseded traditional reporting directly from physicians or pharmaceutical companies (Kuehn 2016). The data gathered in this paper–even though it only represents a subpopulation of those who use social media (Saha et al. 2017b), offers a new lens onto specific groups of people who may have less or more extreme reactions to medications. Including this information in technologies for drug safety monitoring can therefore complement traditional sources, and improve awareness regarding emerging safety issues in a spontaneous fashion —serving as sentinels prompting further exploration in pharmacovigilance research.

Technologies to Support Digital Therapeutics Psychiatrists' view and knowledge of a patient's health is often limited to self-reports and information gathered during in-person therapeutic visits (Vieta 2015). This paper provides a new source of collateral information to support digital therapeutics (Fisher and Appelbaum 2017) and enhance evidence-based, personalized pharmacological treatment. In particular, it reveals the potential to build technologies that

augment information seeking practices of clinicians, e.g., with patient consent, clinicians can learn about the effects and symptomatic expressions shared by patients in the natural course of their lives, and beyond the realms of the therapeutic setting. Further, given the risks posed by prescription drug overdose and abuse (McKenzie and McFarland 2007), increased and finer-grained awareness of the effects of psychiatric medications in specific patients can lead to improved toxicovigilance related interventions.

Policy and Ethics

Despite the potential highlighted above to build novel technologies for regulatory authorities, guidelines on how social media signals should be handled, and their use in the surveillance of the effects of drugs do not yet exist. Although the FDA has released two guidelines on the use of social media for the risk-benefit analysis of prescription drugs (Sarker et al. 2015), they focus on product promotion and "do not establish legally enforceable rights or responsibilities" (FDA 2014). Therefore, the potential (unintended) negative consequences of this work must be considered.

Note that the clinical and technological implications rest upon the names of the medications not being anonymized. We recognize that this surfaces new ethical complexities. For example, while understanding what medications work for which individuals may facilitate "patient-centered" insurance coverage decisions, it can also be (mis)used to decline coverage of specific drugs resulting in "health inequality". Additionally, patients may blindly adopt these findings creating tension in their therapeutic relationship with their clinicians, causing a decrease in medication adherence. We suggest further research investigating and mitigating such potential unintended consequences of the work.

Limitations and Conclusion

We recognize that this study suffers from limitations, and some of these suggest promising directions for future work. Our results on the varied effects of psychiatric medications are likely to be influenced by *selection bias* in those who choose to publicly self-report their medication use on social media. This is especially true given the stigma around mental illness (Corrigan 2004), which is a known obstacle to connecting individuals with mental healthcare. We cannot verify if self-reports of medication use corresponded to their actual use (Ernala et al. 2019). Therefore, the users in our data who chose to self-report their medication usage may represent unique populations with lowered inhibitions. Self-report bias further complicates the types of effects that we observed—different individuals respond differently, as shown in our results, however, our observations are limited to only the types of effects that characterize the individuals in our data. For these same reasons of sampling bias, *we caution against drawing population-wide generalizations* of the effects of psychiatric medication usage.

Despite adopting a causal framework that minimizes confounding effects, *we cannot establish true causality*, and our results are plausibly influenced by the severity in the clinical condition of the individuals. While we considered many confounders in our propensity score matching approach, there are other latent factors that could impact the effects considered here; e.g., duration, history, dosage, and compliance of using self-reported medications; additional medications or adjuvant treatments one might be using. Further, future work can adopt methods such as location-based filtering to better account for geo-cultural and linguistic confounds. Additionally, self-reporting bias about medications can lead to treatment leakage, where some control individuals may be taking medications, but not mentioning it on Twitter.

Our work is not intended as a replacement for clinical trials. In fact, social media lacks many features that clinical trials possess. First, we do not have the notion of a placebo, used to eliminate the confound that simply the perception of receiving a treatment produces non-specific effects. Second, even though we match users based on several characteristics, we do not pre-qualify individuals as potential beneficiaries of a medication. Last, social media analysis does not allow us to closely monitor the treatment, unlike a clinical trial, which results in high variance in the number of measurements that each individual contributes.

Despite corroboration by a psychiatrist, we are limited by what can be observed from an individual's social media data. Without complementary offline information (e.g., the people's physiologies), we cannot ascertain the clinical nature of the mental health outcomes in our data. Further, the symptomatic outcomes themselves, such as measures of depression or suicidal ideation, need additional clinical validation, e.g., based on DSM-5 criteria (APA and others 2013), or the Research Domain Criteria (RDoC) introduced by the National Institutes of Mental Health (Insel et al. 2010). *Without dampening the clinical potentials, we caution against making direct clinical inferences.* Still, while we acknowledge that the medical community rarely adopts the most innovative approaches for immediate use, this work can inspire replication studies in patient populations.

In conclusion, our work represents a novel dynamic viewpoint onto mental health— limitations notwithstanding, it captures the real-time variation and accounts for dynamic systems theory, network theory, and instability mechanisms (Nelson et al. 2017). Such a new window onto the field clearly contrasts the traditional static viewpoint on the effects of psychiatric medications. It warrants further research in this evolving space and opens up interesting opportunities beyond existing reporting methodologies.

Acknowledgement

We thank the members of the Social Dynamics and Well-being Lab at Georgia Tech for their valuable feedback. Saha and De Choudhury were partly supported by NIH grant #R01GM112697. Torous was supported by a patient-oriented research career development award (K23) from NIMH #1K23MH116130-01. Abrahao was supported by a National Natural Science Foundation of China (NSFC) grant #61850410536 and developed part of this research while affiliated with Microsoft Research AI, Redmond.

References

APA, et al. 2013. *Diagnostic and statistical manual of mental disorders, (DSM-5)*. American Psychiatric Pub.

Bagroy, S.; Kumaraguru, P.; and De Choudhury, M. 2017. A social media based index of mental well-being in college campuses. In *Proc. CHI*.

Barchas, J., and Altemus, M. 1999. Monoamine hypotheses of mood disorders. *Basic Neurochemistry*.

Birnbaum, M. L.; Ernala, S. K.; Rizvi, A. F.; De Choudhury, M.; and Kane, J. M. 2017. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J Med Internet Res*.

Blanco, C.; Okuda, M.; Wright, C.; Hasin, D. S.; Grant, B. F.; Liu, S.-M.; and Olfson, M. 2008. Mental health of college students and their non-college-attending peers: results from the national epidemiologic study on alcohol and related conditions. *Arch Gen Psy*.

Bull, S. A.; Hu, X. H.; Hunkeler, E. M.; Lee, J. Y.; Ming, E. E.; Markson, L. E.; and Fireman, B. 2002. Discontinuation of use and switching of antidepressants: influence of patient-physician communication. *Jama*.

Burnap, P.; Colombo, W.; and Scourfield, J. 2015. Machine classification and analysis of suicide-related communication on twitter. In *Proc. ACM conference on hypertext & social media*.

Chancellor, S.; Lin, Z.; Goodman, E. L.; Zerwas, S.; and De Choudhury, M. 2016. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1171–1184. ACM.

Chancellor, S.; Nitzburg, G.; Hu, A.; Zampieri, F.; and De Choudhury, M. 2019. Discovering alternative treatments for opioid use recovery using social media. In *Proc. CHI*.

Chandrasekharan, E.; Samory, M.; Jhaver, S.; Charvat, H.; Bruckman, A.; Lampe, C.; Eisenstein, J.; and Gilbert, E. 2018. The internet's hidden rules: An empirical study of reddit norm violations at micro, meso, and macro scales. *PACM HCI* (CSCW).

Cipriani, A.; Furukawa, T. A.; Salanti, G.; Geddes, J. R.; Higgins, J. P.; Churchill, R.; Watanabe, N.; Nakagawa, A.; Omori, I. M.; McGuire, H.; et al. 2009. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The lancet* 373(9665):746–758.

Cipriani, A.; Furukawa, T. A.; Salanti, G.; Chaimani, A.; Atkinson, L. Z.; Ogawa, Y.; Leucht, S.; Ruhe, H. G.; Turner, E. H.; Higgins, J. P.; et al. 2018. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet* 391(10128):1357–1366.

Coppersmith, G.; Harman, C.; and Dredze, M. 2014. Measuring post traumatic stress disorder in twitter. In *ICWSM*.

Corrigan, P. 2004. How stigma interferes with mental health care. *American Psychologist* 59(7):614.

Corsello, S. M.; Bittker, J. A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J. E.; Johnston, S. E.; Vrcic, A.; Wong, B.; Khan, M.; et al. 2017. The drug repurposing hub: a next-generation drug library and information resource. *Nature medicine* 23(4):405.

Coupland, C.; Dhiman, P.; Morriss, R.; Arthur, A.; Barton, G.; and Hippisley-Cox, J. 2011. Antidepressant use and risk of adverse outcomes in older people: population based cohort study. *Bmj*.

De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting depression via social media. In *ICWSM*.

De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; and Kumar, M. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *CHI*.

Dos Reis, V. L., and Culotta, A. 2015. Using matched samples to estimate the effects of exercise on mental health from twitter.

Dubovsky, S. L. 2018. What is new about new antidepressants? *Psychotherapy and psychosomatics*.

Dutta, S.; Ma, J.; and De Choudhury, M. 2018. Measuring the impact of anxiety on online social interactions. In *ICWSM*.

Ernala, S. K.; Rizvi, A. F.; Birnbaum, M. L.; Kane, J. M.; and De Choudhury, M. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *CSCW*.

Ernala, S. K.; Birnbaum, M. L.; Candan, K. A.; Rizvi, A. F.; Sterling, W. A.; Kane, J. M.; and De Choudhury, M. 2019. Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. In *ACM CHI*.

FDA, et al. 2014. Guidance for industry: internet/social media platforms with character space limitations; presenting risk and benefit information for prescription drugs and medical devices.

Fisher, C. E., and Appelbaum, P. S. 2017. Beyond googling: The ethics of using patients' electronic footprints in psychiatric practice. *Harvard review of psychiatry* 25(4):170–179.

Frommer, D. A.; Kulig, K. W.; Marx, J. A.; and Rumack, B. 1987. Tricyclic antidepressant overdose: a review. *Jama*.

Gaur, M.; Kursuncu, U.; Alambo, A.; Sheth, A.; Daniulaityte, R.; Thirunarayan, K.; and Pathak, J. 2018. Let me tell you about your mental health!: Contextualized classification of reddit posts to dsm-5 for web-based intervention. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 753–762. ACM.

Geddes, J.; Freemantle, N.; Mason, J.; Eccles, M.; and Boynton, J. 2000. Selective serotonin reuptake inhibitors (ssris) versus other antidepressants for depression. *Cochrane Database Syst Rev*.

Gillman, P. K. 2007. Tricyclic antidepressant pharmacology and therapeutic drug interactions updated. *Br. J. Pharmacol.*

Gohol, J. M. 2018. `https://psychcentral.com/blog/top-25-psychiatric-medications-for-2016/`. Accessed: 2018-09-08.

Hannan, E. L. 2008. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC: Cardiovascular Interventions* 1(3):211–217.

Harpaz, R.; DuMouchel, W.; Schuemie, M. J.; Bodenreider, O.; Friedman, C.; Horvitz, E.; et al. 2017. Toward multimodal signal detection of adverse drug reactions. *J. Biomed. Inform.*

Hasin, D. S.; Goodwin, R. D.; Stinson, F. S.; and Grant, B. F. 2005. Epidemiology of major depressive disorder: results from the national epidemiologic survey on alcoholism and related conditions. *Arch. Gen. Psychiatry*.

Imbens, G. W., and Rubin, D. B. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge.

Insel, T.; Cuthbert, B.; Garvey, M.; Heinssen, R.; Pine, D. S.; Quinn, K.; Sanislow, C.; and Wang, P. 2010. Research domain criteria (rdoc): toward a new classification framework for research on mental disorders.

Kerr, G.; McGuffie, A.; and Wilkie, S. 2001. Tricyclic antidepressant overdose: a review. *Emergency Medicine Journal*.

Kiciman, E.; Counts, S.; and Gasser, M. 2018. Using longitudinal social media analysis to understand the effects of early college alcohol use. In *ICWSM*.

Klein, A.; Sarker, A.; Rouhizadeh, M.; O'Connor, K.; and Gonzalez, G. 2017. Detecting personal medication intake in twitter: An annotated corpus and baseline classification system. *BioNLP 2017*.

Kuehn, B. M. 2016. Fdaś foray into big data still maturing. *Jama*.

Lamont, A.; Lyons, M. D.; Jaki, T.; Stuart, E.; Feaster, D. J.; et al. 2018. Identification of predicted individual treatment effects in randomized clinical trials. *Stat. Methods Med. Res*.

Lardon, J.; Abdellaoui, R.; Bellet, F.; Asfari, H.; Souvignet, J.; Texier, N.; Jaulent, M.-C.; Beyens, M.-N.; Burgun, A.; and Bousquet, C. 2015. Adverse drug reaction identification and extraction in social media: A scoping review. *J. Med. Internet Res*.

Lexchin, J.; Bero, L. A.; Djulbegovic, B.; and Clark, O. 2003. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *Bmj* 326(7400):1167–1170.

Lin, H.; Jia, J.; Guo, Q.; Xue, Y.; Li, Q.; Huang, J.; Cai, L.; and Feng, L. 2014. User-level psychological stress detection from social media using deep neural network. In *Proc. ACM Multimedia*.

Liu, J.; Weitzman, E. R.; and Chunara, R. 2017. Assessing behavioral stages from social media data. In *CSCW*.

Lopez-Munoz, F., and Alamo, C. 2009. Monoaminergic neurotransmission: The history of the discovery of antidepressants from 1950s until today. *Current Pharmaceutical Design*.

MacLean, D.; Gupta, S.; Lembke, A.; Manning, C.; and Heer, J. 2015. Forum77: An analysis of an online health forum dedicated to addiction recovery. In *CSCW*.

McKenzie, M. S., and McFarland, B. H. 2007. Trends in antidepressant overdoses. *Pharmacoepidemiology and drug safety*.

Moncrieff, J., and Cohen, D. 2009. How do psychiatric drugs work? *BMJ* 338:1535.

Nelson, B.; McGorry, P. D.; Wichers, M.; Wigman, J. T.; and Hartmann, J. A. 2017. Moving from static to dynamic models of the onset of mental disorder: a review. *JAMA psychiatry*.

Nikfarjam, A.; Sarker, A.; Oconnor, K.; Ginn, R.; and Gonzalez, G. 2015. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J. Am. Med. Inform.*

Oliver, D.; Davies, C.; Crossland, G.; Lim, S.; Gifford, G.; McGuire, P.; and Fusar-Poli, P. 2018. Can we reduce the duration of untreated psychosis? a systematic review and meta-analysis of controlled interventional studies. *Schizophrenia bulletin*.

Olteanu, A.; Varol, O.; and Kiciman, E. 2017. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proc. CSCW*.

Pavalanathan, U., and Eisenstein, J. 2016. More emojis, less:) the competition for paralinguistic function in microblog writing.

Pennebaker, J. W.; Mehl, M. R.; and Niederhoffer, K. G. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54(1):547–577.

Powell, T. R.; Murphy, T.; Lee, S. H.; Price, J.; Thuret, S.; and Breen, G. 2017. Transcriptomic profiling of human hippocampal progenitor cells treated with antidepressants and its application in drug repositioning. *Journal of Psychopharmacology*.

Rosenblat, J. D.; Kakar, R.; and McIntyre, R. S. 2016. The cognitive effects of antidepressants in major depressive disorder: a systematic review and meta-analysis of randomized clinical trials. *International Journal of Neuropsychopharmacology* 19(2).

Saha, K., and De Choudhury, M. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proc. ACM HCI (CSCW)*.

Saha, K.; Chan, L.; De Barbaro, K.; Abowd, G. D.; and De Choudhury, M. 2017a. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proc. ACM IMWUT*.

Saha, K.; Weber, I.; Birnbaum, M. L.; and De Choudhury, M. 2017b. Characterizing awareness of schizophrenia among facebook users by leveraging facebook advertisement estimates. *Journal of medical Internet research* 19(5):e156.

Saha, K.; Bayraktaraglu, A. E.; Campbell, A. T.; Chawla, N. V.; De Choudhury, M.; D'Mello, S. K.; Dey, A. K.; et al. 2019a. Social media as a passive sensor in longitudinal studies of human behavior and wellbeing. In *CHI Ext. Abstracts*. ACM.

Saha, K.; Torous, J.; Ernala, S. K.; Rizuto, C.; Stafford, A.; and De Choudhury, M. 2019b. A computational study of mental health awareness campaigns on social media. *Translational behavioral medicine*.

Saha, K.; Weber, I.; and De Choudhury, M. 2018. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. In *ICWSM*.

Sarker, A.; Ginn, R.; Nikfarjam, A.; OConnor, K.; Smith, K.; Jayaraman, S.; Upadhaya, T.; and Gonzalez, G. 2015. Utilizing social media data for pharmacovigilance: a review. *J. Biomed. Inform.*

Scripts, E. 2018. `http://lab.express-scripts.com/lab/drug-trend-report/~/media/29f13dee4e7842d6881b7e034fc0916a.ashx`.

Shippee, N. D.; Shah, N. D.; May, C. R.; Mair, F. S.; and Montori, V. M. 2012. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. *Journal of clinical epidemiology* 65(10):1041–1051.

Szegedi, A.; Kohnen, R.; Dienel, A.; and Kieser, M. 2005. Acute treatment of moderate to severe depression with hypericum extract ws 5570 (st john's wort): randomised controlled double blind non-inferiority trial versus paroxetine. *Bmj* 330(7490):503.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

Trivedi, M.; Rush, A.; Wisniewski, S.; et al. 2006. Evaluation of outcomes with citalopram for depression using measurement-based care in star* d: implications for clinical practice. *Am. J. Psychiatry*.

Vieta, E. 2015. Personalised medicine applied to mental health: precision psychiatry. *Revista de psiquiatria y salud mental*.

White, R. W.; Wang, S.; Pant, A.; Harpaz, R.; Shukla, P.; Sun, W.; DuMouchel, W.; and Horvitz, E. 2016. Early identification of adverse drug reactions from search log data. *J. Biomed. Inform.*

WHO. 2002. The importance of pharmacovigilance. *WHO*.

Yoo, D. W., and De Choudhury, M. 2019. Designing dashboard for campus stakeholders to support college student mental health. In *Pervasive Health*.

LibRA: On LinkedIn based Role Ambiguity and Its Relationship with Wellbeing and Job Performance

KOUSTUV SAHA, Georgia Institute of Technology, USA
MANIKANTA D. REDDY, Georgia Institute of Technology, USA
STEPHEN M. MATTINGLY, University of Notre Dame, USA
EDWARD MOSKAL, University of Notre Dame, USA
ANUSHA SIRIGIRI, Dartmouth College, USA
MUNMUN DE CHOUDHURY, Georgia Institute of Technology, USA

Job roles serve as a boundary between an employee and an organization, and are often considered building blocks in understanding the behavior and functioning of organizational systems. However, a lack of clarity about one's role, that is, one's work responsibilities and degree of authority, can lead to absenteeism, turnover, dissatisfaction, stress, and lower workplace performance. This paper proposes a methodology to quantitatively estimate role ambiguity via unobtrusively gathered data from LinkedIn, shared voluntarily by a cohort of information workers spanning multiple organizations. After successfully validating this LinkedIn based measure of Role Ambiguity, or `LibRA` against a state-of-the-art gold standard, drawing upon theories in organizational psychology, we examine the efficacy and convergent validity of `LibRA` in explaining established relationships of role ambiguity with wellbeing and performance measures of individuals. We find that greater `LibRA` is associated with depleted wellbeing, such as increased heart rate, increased arousal, decreased sleep, and higher stress. In addition, greater `LibRA` is associated with lower job performance such as decreased organizational citizenship behavior and decreased individual task performance. We discuss how `LibRA` can help fill gaps in state-of-the-art assessments of role ambiguity, and the potential of this measure in building novel technology-mediated strategies to combat role ambiguity in organizations.

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

Additional Key Words and Phrases: LinkedIn; role ambiguity; social media; wellbeing; passive sensing; job performance; productivity; stress

ACM Reference Format:
Koustuv Saha, Manikanta D. Reddy, Stephen M. Mattingly, Edward Moskal, Anusha Sirigiri, and Munmun De Choudhury. 2019. LibRA: On LinkedIn based Role Ambiguity and Its Relationship with Wellbeing and Job Performance. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 137 (November 2019), 30 pages. https://doi.org/10.1145/3359239

1 INTRODUCTION

Employee job satisfaction is of prime interest to both individuals as well as organizations. The complexities related to an individual's job role, or the *expectations applied to an individual within and beyond an organization's boundaries* can impact their job satisfaction [153]. In fact, any sort of

Authors' addresses: Koustuv Saha, koustuv.saha@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA; Manikanta D. Reddy, mani@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA; Stephen M. Mattingly, smattin1@nd.edu, University of Notre Dame, South Bend, Indiana, USA; Edward Moskal, emoskal@nd.edu, University of Notre Dame, South Bend, Indiana, USA; Anusha Sirigiri, anusha.sirigiri@dartmouth.edu, Dartmouth College, Hanover, New Hampshire, USA; Munmun De Choudhury, munmund@gatech.edu, Georgia Institute of Technology, Atlanta, Georgia, USA.

2573-0142/2019/11-ART137 $15.00
https://doi.org/10.1145/3359239

discrepancy between *what* an employer expects and *what* an employee does at the workplace can impact wellbeing and performance, as employees can find themselves pulled in various directions as they try to respond to the many statuses they hold. According to the "Role Theory", role conflict, role ambiguity, and role overload are three aspects of job role that contribute to workplace stress, or the stress that arises if the demands of an individual's roles and responsibilities exceed their capacity and capability to cope [79, 112]. Among the role constructs, role ambiguity has been considered to be the most significant one, and it is also the focus of the current paper [79].

Role ambiguity is broadly considered to include uncertainties about role definition, expectations, responsibilities, tasks, and behaviors involved in one or more facets of the task environment [72, 79, 127]. Role ambiguity has both objective and subjective components — Objective role ambiguity refers to external conditions in the individual's workplace environment, whereas subjective role ambiguity relates to the amount of ambiguity that the individual perceives in their workplace owing to the information gap that they face [79]. Further, role ambiguity leads to consequences related to dissatisfaction, distrust, lack of loyalty, turnover, absenteeism, low performance, anxiety-stress, and increased heart rate [153]. There is sufficient evidence demonstrating how role ambiguity negatively affects one's organizational life in terms of their physiological, behavioral, psychological, and performance related measures [78, 127].

Traditionally, role ambiguity is measured using survey instruments that record employee responses to their perceived clarity of assigned tasks, expectations on the job, expectations of peers, and if these peers explicitly mention their expectations from the focal employee [117]. In particular, these methods not only suffer from subjective biases [136], but also are only able to capture the "perceived" component of role ambiguity. Individuals may or may not be aware that they are working on things beyond their job requirements, such as when there is an information gap, or if they are investing their effort to gain knowledge and experience [47, 81]. Thus, it is unclear how useful these measures are [116], and researchers have argued that the lack of an instrument capable of measuring objective and perceived facets of ambiguity may have impeded both theory development and application of research results [16].

Further, with the development and adoption of technology in several sectors of the workplace, the landscape of work is evolving at an unprecedented speed. This also demands continuous skill development [24, 75]; a recent study by McKinsey Global Institute predicts enormous workforce transitions in the years ahead, estimating by 2030, as many as 375M workers globally will likely need to transition to new occupational categories and learn new skills [93]. However, there is no defined approach for organizations to proactively gauge individuals' fit with their assigned roles, no guidance for interventions to help them overcome role ambiguity. An organization that can proactively deal with role ambiguity will benefit from employees with increased satisfaction, wellbeing, and productivity in general.

Our study contributes to the above research gap and advances the theory by introducing a novel way of measuring role ambiguity. To the best of our knowledge, our study is the first to empirically and objectively measure role ambiguity via LinkedIn, a professional social networking platform where career profiles are publicly shared by employees with self-descriptions of their job titles and role descriptions. Juxtaposing traditional surveys with modern sensor derived measures of wellbeing, we combine methods adopted from natural language analysis and statistical modeling to examine the relationship of LinkedIn based role ambiguity (`LibRA`) with the wellbeing and job performance of individuals — the two important facets corresponding to one's job satisfaction [127].

Aim 1. To measure role ambiguity using unobtrusively obtained LinkedIn data.
Aim 2. To examine the relationship of Linkedin based Role Ambiguity (`LibRA`) with individual wellbeing and job performance.
Aim 3. To investigate what factors may contribute to one's `LibRA`, relating to their intrinsic traits, LinkedIn's platform-specific characteristics, and preferences and goals of use of professional social networking service.

For our first research aim, we model LinkedIn based Role Ambiguity (LibRA) as a lexico-semantic difference between the job description of an individual's role as self-portrayed on their Linkedin profile and what is posted by the company for that particular role. To compute this difference, we first employ natural language analysis techniques of word-embeddings to obtain the vector representations of job descriptions along eight facets of job role, namely *abilities, interests, knowledge, skills, work activities, work context, work styles, and work values* [141].

In our second research aim, in a theoretically grounded fashion, we test for hypotheses that examine the relationship of LibRA with an individual's 1) wellbeing related measures, namely their heart rate, arousal, sleep, and work-hours, and 2) job performance related measures, namely their individual task performance, in-role behavior, and organizational citizenship behavior. Our findings align with the propositions put forth by role theory, that greater LibRA measure is associated with factors that are related to depleted wellbeing such as, increased heart rate, increased arousal, decreased sleep, and decreased work hours, and is associated with lower job performance such as decreased task performance and decreased organizational citizenship behavior.

Finally, in our third research aim, we reflect back to investigate what factors contribute to LibRA as a measure. We contextualize one's self-presentation behavior on LinkedIn, by situating our observations in the literature on social sciences, social computing, and organizational studies. Our observations make valuable insights into the unobservable and unaccounted factors, such as an individual's mindset, job-related motivation, and platform-related nuances.

We discuss the theoretical, practical, and design implications that surround this new measure of role ambiguity assessed from people's professional social networking data, from the perspective of employees, organizations, and social computing platforms. Our research contributes to the growing interest on the topic of "Future of Work at the Human-Technology Frontier"[1], wherein we present new technology-facilitated means to improve workplace "health", performance, and functioning.

Privacy, Ethics, and Disclosure. This work is committed to secure the privacy of the employees and their employers, whose data on individual difference attributes are used. These individuals signed informed consent to provide the survey responses as a part of the Tesserae study, which was approved by the relevant Institutional Review Boards at researcher institutions. In addition, despite working with publicly posted job descriptions on websites such as LinkedIn, Glassdoor, Indeed, and company websites, this paper anonymizes each of these job titles and job descriptions. Finally, this paper reports paraphrased excerpts of LinkedIn self-described portfolios of the individuals. Together, these measures balance the sensitivity of privacy, traceability, and identifiability, as well as provide a context in readership. Even accounting for the benefits, we recognize and acknowledge the limitations of our methodological approach, and the potential misuses, risks, and ethical consequences involved with this kind of research, which we elaborate in Section 7.

2 BACKGROUND, THEORETICAL UNDERPINNINGS, AND RESEARCH AIMS

2.1 Role Ambiguity in Organizational Psychology

The theory of organizational role dynamics outlines role conflict, role ambiguity, and role overload as aspects of the job role that contribute to workplace stress [11, 112, 116]. Organizational psychology literature emphasized these characteristics of job role for at least five decades now and ample empirical evidence exists as to how role ambiguity impacts psychological wellbeing and productivity of individuals [79]. Kahn et al. defined role ambiguity as "*the discrepancy between the information available to the person and that which is required for adequate performance of his role*".

Over the years, surveys have been used to measure role constructs, and there has been no consensus on the quality of the measure [116, 134]. In surveys, *role conflict* is measured by asking individuals if they had to cater to addressing multiple co-worker needs simultaneously, and if they have to break rules to get things done. *Role ambiguity* is measured using responses to whether there

[1] *https://www.nsf.gov/eng/futureofwork.jsp*

are defined objectives for the role and if they can be well predicted by employees. *Role overload* is measured by the perception of employees, if they can get work done in available time and if a lot more is expected from them [112]. Among these, role ambiguity is prone to more significant changes and is more dynamic owing to the technological evolution in the industry. According to the role theory, role ambiguity is associated with the increase in the likelihood that the individual would be dissatisfied with their role, would experience anxiety, and would thus perform less effectively [117].

Since role ambiguity revolves around the expectations surrounding one's job role, this paper leverages an unobtrusive data source (self-described / self-presented) professional portfolio on social media (LinkedIn) to infer their role ambiguity. We validate our measure of role ambiguity on the basis of theoretically driven hypotheses, grounded in the literature on the relationship of role ambiguity with wellbeing and performance, which we discuss in the following two subsections.

2.1.1 Role Ambiguity and Wellbeing. Research has demonstrated that role ambiguity has negative consequences on employee wellbeing [134]. Role ambiguity is one of the antecedents of job satisfaction, and an increase in job satisfaction leads to lower stress [139] and higher intrinsic motivation of employees [158]. When employees are intrinsically motivated, they tend to have lesser somatic symptoms and lower anxiety [90]. In tune with the traditional principal-agent problem, lower intrinsic motivation leads to a lesser effort expended by employees at work [46].

While there is no single conceptualization of wellbeing, the broad categories that wellbeing encompasses are physiological, psychological and behavioral aspects [78, 127]. Physiological indicators include factors such as blood pressure, heart conditions, and general physical health. Psychological indicators include affect, frustration, anxiety, stress, and arousal. Behavioral aspects include those that an employee has a choice to make, like the time spent at work, the time taken for breaks during work, mobility to another employment (turnover), hours of sleep, etc.

Within the scope of our dataset, we study the relationship of `LibRA` with one's physiological measures (heart rate and sleep [21, 25]), psychological measures (stressful arousal [139]), and behavior at the workplace (time spent at desk and time spent at workplace [162]). Specifically, we test for the following hypotheses in the relationship of `LibRA` with wellbeing attributes.

H1. Greater role ambiguity is associated with increased heart rate.
H2. Greater role ambiguity is associated with increased arousal.
H3. Greater role ambiguity is associated with decreased sleep.
H4. Greater role ambiguity is associated with reduced work-hours.

2.1.2 Role Ambiguity and Job Performance. While job satisfaction is an important antecedent of job performance, the others include employee motivation and engagement. A more comprehensive explanation of job performance alludes to taking into account intrinsic motivation and employee engagement [77, 115, 125], which are shown to be determinants of effort exerted by employees. Such discretionary effort by employees improves their engagement with the organization [97]. Employee engagement has been defined in many ways [92], with satisfaction and motivation being core components. In fact, motivation has been perceived as employee engagement in a number of prior work [130, 156]. Yun et al. provide evidence that role ambiguity moderates the relationship between self-enhancement motive and job performance of an individual [160].

Role ambiguity consists of the uncertainty regarding tasks that an employee needs to perform as part of their job role in the company. An employee with greater clarity will be able to better perform the required tasks. One plausible mechanism that can explain this higher performance is the intrinsic motivation of an employee [90, 117]. Lower role ambiguity or higher role clarity makes it easier to meet the expectations, the employee more motivated and such intrinsically motivated employees perform better and more efficient [46, 48]. Employees with higher job satisfaction are intrinsically motivated and strive harder at work which contributes to their performance. Thereby,

the exposure to role stressors (such as role ambiguity) affects an individual's capacity to control their work environment, which in turn adversely affects their ability to function effectively [88, 99].

Within the scope of our dataset, we study the relationship of LibRA with two dimensions of job performance [118, 154, 157] — 1) task performance and 2) organizational citizenship behavior. Here, *task performance* is a combination of individual task proficiency (three-item scale) [58] and in-role behavior (seven-item scale) [157]. Both of them measure the ability of an individual to adequately execute their assigned duties, and their proficiency at performing activities that drive an organization's *technical core*, or production processes that drive the conversion of input into output [143]. On the other hand, *organizational citizenship behavior(s)* is a set of pro-organizational actions that are not formally rewarded but demonstrate how an individual contributes to welfare and effectiveness within the organization [106, 111, 128]. Examples include helping other co-workers, kind gestures like volunteering in activities that are not part of work, or strictly adhering to rules at the expense of personal convenience. These are subjective and self-reported/assessed measures of job performance and we do not use any objective measures or supervisor/peer- assessments like ratings, evaluations, sales or profits, which are more likely to suffer from leniency, halo error, criterion contamination and deficiency [65, 70]. Prior literature in organizational behavior dominantly uses these subjective measures and we rely on the extant literature for the validity of these measures [154]. We test the following hypotheses for LibRA with respect to job performance.

H5. Greater role ambiguity is associated with decreased task performance.
H6. Greater role ambiguity is associated with decreased organizational citizenship.

2.2 Social Media Technologies and Workplace Behavior

In the last decade, researchers have used social media technologies to understand employee behavior at workplaces [33]. In their seminal work, Ehrlich and Shami compared employees' use of social media platforms inside and outside the workplace, particularly their motivations in their use of social media, particularly Twitter [40]. They report that social media use (both at home and work) made workers, especially mobile workers, feel more connected to other employees, and provided an avenue to boost personal reputation at the workplace. Studies also found that social media use is positively correlated workplace wellbeing [131]. Increased social media interactions within the workplace, through platforms such as IBM's Beehive, were found to improve both personal and professional networking, career advancements, and innovation [36, 37, 42, 50]. Other works find positive relationships between workplace and employee behavior, such as wellbeing, experiences, and engagement through social media technologies [33, 41, 44, 61]. In an early work, Skeels and Grudin conducted a longitudinal study of the motivations and use of social media platforms by workplace employees [135]. Taken together, social media use, both in and outside of the workplace contribute to the wellbeing and professional benefits through increased connectivity, reputation building, and networking opportunities.

In the same body of research, social media and online engagement platforms have facilitated an effective means to study employee behavior and satisfaction — a body of research that is extensive in CSCW and HCI area [5, 33, 105, 130, 132, 135]. A variety of analytical and computational approaches on language and network dynamics have been applied to glean correlates of employee job satisfaction and wellbeing, such as engagement [66, 105, 130], employee affect [33, 121], social pulse [131], reputation [74], organizational relationships [17, 52, 104], and workplace behavior [94]. Lee and Kang used Glassdoor data to study the influence job satisfaction factors, and their influence on employee retention and turnover [89]. These studies indicate the value of such unobtrusive data sources in understanding workplace experiences.

In the professional networking space, LinkedIn has emerged as the primary social media platform [140, 148]. This platform, which was initially viewed as a "repository of webpages", gradually evolved to be informally known as "Facebook in a suit" [5, 151]. LinkedIn allows the individuals to self-describe and self-promote their professional portfolio to either seek for new jobs, or to use it

as their professional networking and webpage. Guillory and Hancock found that the public-facing nature of LinkedIn influences an individual's accountability and reduces deception in their self-description of their professional portfolio, which also aligns with Donath et al.'s early research on identity and deception in online spaces [38]. Researchers have studied the differences and similarities in the self-presentation behavior and use of LinkedIn in comparison to personal social media platforms such as Facebook and Twitter [5, 135, 151, 163]. In fact, organizations' use of LinkedIn has grown tremendously over the years, which also implicitly puts peer- and societal- pressure on individuals to own and maintain LinkedIn accounts [84]. Utz and Breuer recently studied the individual-specific factors that influence their behavior on LinkedIn in terms of networking and informational benefits that the platform facilitates [149], van de Ven et al. inferred personality traits on LinkedIn self-presentations of individuals, and Zide et al. studied how LinkedIn profiles differ across occupations. Zhang et al. studied employees' privacy perceptions on social media [163].

While all these sources are combined under a broad umbrella of 'social media', the motivations to use any of these platforms might differ based on individual and platform-specific characteristics [5, 33, 40, 120, 163]. On LinkedIn, the primary motivation to use is to gain professional visibility [5, 20, 85]. This professional visibility might be used by individuals to find jobs or switch jobs, or escalate to a better position within the same job, or to reach out to a broader community in general (like students in case the focal individual is in academia, or venture capitalists and funding agencies if the individual is in top management teams of start-ups) [5, 135]. Since preparing this profile is an exercise that makes the individual reflect upon their work, it also motivates to work towards building a better profile or developing new skills that would enhance their profile. While such motivation [49] leads individuals to work towards their personal goals, these goals might not align with organizational goals and this friction might contribute to stress leading to lower employee wellbeing, or can improve productivity because the employee works harder [161].

Our third research aim draws upon the literature in organizational studies, along with the literature to contextualize and discuss the unaccounted factors that may affect one's activity and self-presentation behavior on social media, and in turn, influence our measure of `LibRA`. Further, building on this body of work, our study leverages LinkedIn data to infer role ambiguity, a role-construct drawn on organizational psychology research, and then validates theoretically grounded hypotheses of individual wellbeing and performance. In doing so, we extend the CSCW community's long-drawn interest in understanding the role of technology in the workplace and on work (e.g., [110]) by investigating (social computing) platform, individual, and organization-specific complexities that should be considered to reliably and practically implement a measure like `LibRA`.

3 STUDY AND DATA

3.1 Study: The Tesserae Project

Our dataset comes from a large-scale multi-sensor study of workplace behaviors, called the Tesserae Project [98, 103]. This study, that was approved by the Institutional Review Board (IRB) at the researchers' institutions, recruited 757 participants[2] who are information workers in cognitively demanding fields (e.g. engineers, consultants, managers) across the United States. These participants who were recruited from January 2018 through July 2018, completed an initial set of questionnaires related to demographics, job performance, personal attributes, and wellbeing, administrated via psychometrically validated survey instruments, as well as received daily surveys on a set of these attributes. Participants also received three sensors: location-tracking Bluetooth beacons; 2) a wearable; 3) a phone agent—a smartphone application [155]. In addition, some participants authorized collection of their historical social media data. As compensation, participants either received a series of staggered stipends totaling up to $750 or they participated in a set of weekly

[2]Note that this is still an ongoing study and this paper uses the passively sensed data collected until November 13th, 2018. A random sample of 154 participants in our study was "blinded at source" to the researchers, and their data is put aside only for external validation at the end of the study period.

lottery drawings (multiples of $250 drawings) depending on their employer restrictions. Because the participants were enrolled over a 6 month period of time (January to July 2018) in a staggered fashion, data collection varied with a range of time between 59 days and 97 days (68 days on an average) (see Fig. 1(a) for the distribution).

Participant Privacy and Consent. Given the sensitive nature of the data being collected, participant privacy was a key concern in the study. The participants were provided with an informed-consent describing each sensing stream, and technical specifications listed what each device was capturing and how it would be secured and stored. The participants needed to consent to each sensing stream individually, and they had the provision to clarify their queries / concerns about the sensing streams, and they could opt out of any of them [98]. Their data was de-identified and stored in secured databases and servers which were physically located in one of the researcher institutions, and had limited access privileges. Participants were made aware that they could voluntarily drop out via an email at any point during the year-long study period. Participants could also specifically request their data deletion from the database.

3.2 Social Media (LinkedIn) Data

Participants authorized access to their social media data through an Open Authentication (OAuth) based data collection infrastructure that was developed in-house [119, 121]. In particular, we asked permission from participants to provide their Facebook and LinkedIn data, *unless they opted out, or did not have either of these accounts*. Note that we asked consent from only those participants who had existing Facebook or LinkedIn accounts from before the study. The participants could also optionally authorize their Twitter, Instagram, GMail, and Calendar data.

Out of the 757 participants in the study, 529 provided their LinkedIn data. Our work accounts for those with self-described portfolios and their passively sensed and self-reported wellbeing and job performance data. Therefore, we filter out "blinded" participants and those without any self-description in their LinkedIn profile, particularly in their profile and job summary, leading us to a LinkedIn dataset of 257 individuals — all the ensuing analyses in this paper is limited to these 257 individuals' data. Corresponding to every participant, we obtained their self-presented profile and job summary which includes current and previous jobs. Fig. 1(c) shows the top job titles in our dataset, and Fig. 1(d&e) shows two word-trees of profile summaries on two top representative keywords ("professional" and "skill") in our dataset. These word-trees give a sense of how individuals self-present their job summaries on their LinkedIn profiles; for example, within skills, we find occurrences of both tangible/technical skills (eg. *sap, technology, sales, microsoft office*) and intangible/people skills (eg. *leadership, communication, analytical, interpersonal*).

3.3 Self-Reported Data

As mentioned above, the enrollment process consisted of responding to a set of initial survey questionnaires related to demographics (age, sex, education, type of occupation, role in the company, and income). Participants were additionally required to answer an initial ground-truth battery, a set of survey questionnaires that measured their self-reported assessments of personality traits and executive function. Throughout their study period, they received daily or periodic validated surveys that recorded their self-reported assessments of job performance.

Fig. 1b plots the demographic distribution in this dataset. The 257 participants with *complete* LinkedIn data consist of 150 males and 107 females. The average age of the participants is 35.2 years (stdev. = 9.5). These participants belong to 60 unique companies, and among these, the top three companies, 103 belong to C_1 (a large-scale multinational firm), 54 belong to C_2 (a mid-size product-centric firm), and 17 belong to C_3 (a research organization). In terms of job role, the data contains 128 supervisors and 139 non-supervisors. In job sector, 202 participants belong to Non-IT sector, and 55 participants belong to the IT sector. In terms of tenure, while a majority of the individuals (53) have been at their current organization for over eight years, 113 individuals have

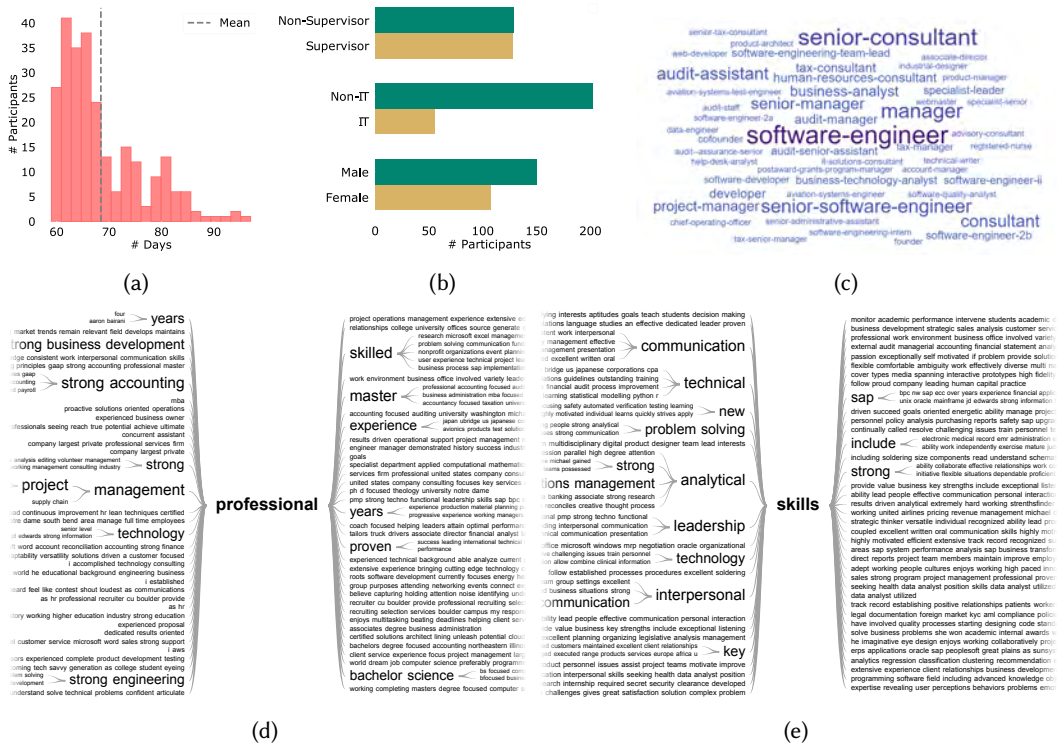


Fig. 1. Distribution in the dataset on (a) study period per participant, (b) demographic and job-role based characteristics, (c) word-cloud on the job roles on LinkedIn data, (d&e) Word tree visualizations on two top-occurring keywords (professional and skills) in the LinkedIn profile descriptions: These visualizations represent the content in the form of co-occurrences of keywords in the dataset. The font size of keywords are proportional to their occurrence along with surrounding co-occurring keywords. For example, management profesional, technology professional, professional keywords, analytical skills, technological skills, etc.

been at their current organization between three to eight years, and 101 individuals have been at their current organization for less than three years. For education, most participants have a college (52%) or master's degree (35%). In income, the participants are more evenly distributed, with the majority (64%) of the participants similarly distributed in the 50K-75K, 75K-100K, and more than 150K USD income brackets.

Next, we discuss their self-assessed individual difference attributes. We obtained the participants' big-five **personality traits**, as assessed by the Big Five Inventory (BFI-2) scale [138, 142], and **executive function**, especially their fluid and crystallized intelligence, as assessed by the Shipley scale [23, 129, 133]. For personality traits, our dataset shows a mean openness of 3.86 (std. = 0.54), mean conscientiousness of 3.87 (std. = 0.66), mean extraversion of 3.41 (std. = 0.67), mean agreeableness of 3.85 (std. = 0.55), and mean neuroticism of 2.51 (std. = 0.77). For executive function, our dataset shows a mean fluid intelligence of 33.38 (std. = 4.18), and mean crystallized intelligence of 16.91 (std. = 2.79).

For job performance, we obtain two kinds of measures:

Task Performance. To assess task performance, we use two scales, IRB (In-Role-Behavior) [157] and ITP (Individual Task Proficiency) [58]. The IRB scale contains seven items including questions such as *adequately performed assigned duties, failed to perform essential duties, performed expected*

tasks, etc., each of which can be rated on a scale of 1 (strongly disagree) to 7 (strongly agree). On the other hand, the ITP scale contains three items, *carried out core parts of the job well, completed core tasks well using standard procedures, and ensured that the tasks were completed properly*, each of which can be rated on a scale of 1 (very little) to 5 (a great deal). Together, these instruments measure an individual's ability to adequately execute their assigned duties, and their proficiency at performing activities that drive an organization's technical core [14, 154].

Organizational Citizenship Behavior. We administer the OCB scale to measure organizational citizenship behavior [45]. Organizational citizenship behaviors characterize an individuals activities that are not typically or formally rewarded by the management, or voluntary activities outside one's core responsibilities, but which promote the welfare and effectiveness of the organization and its members [26, 111]. The survey instrument contains eight items, each of which asks the participant to self-reflect (yes/no), if they, *went out of their way to be a good employee, were respectful of other people's needs, displayed loyalty to my organization, praised or encouraged someone*, etc.

3.4 Passively Sensed Behavior and Wellbeing Data

To passively sense participants' behavior and wellbeing measures of participants, the study deployed three modalities of sensing technologies [12, 98] — 1) **bluetooth beacons** were provided to the participants (two static and two portable Gimbal beacons [4]) to essentially sense their presence at work and home locations, and consequently to help assess their commute and desk time as well, 2) A **wearable** (Garmin Vivosmart [3]) was provided to each participant to continually track their health measures, such as heart rate, arousal, and physical activity in the form of sleep, foot steps, and calories lost, and 3) A **smartphone application** was installed on the participants' smartphones to leverage their smartphone based mobile sensors to track their mobility and physical activity [155].

4 MEASURING ROLE AMBIGUITY FROM LINKEDIN (LibRA)

Why LinkedIn? LinkedIn is a professional social networking platform (launched in 2003) that allows individuals to create and publish their professional profiles and describe and their portfolios. Although LinkedIn is biased towards individuals' positive self-presentation and self-promotion, the non-anonymity and public-facing nature of the platform also influences individuals to be less deceptive and more accountable in their profiles [60]. In line with Goffman's theory of self-presentation, LinkedIn provides an ideal platform for individuals to present their "professional" selves to the online audience [54, 151]. Because LinkedIn is a non-anonymous platform, where individual identity (at least the name) is disclosed, it somewhat helps promote trust and accountability on the platform [38]. Therefore, it suits the choice of our dataset where we seek to obtain self-presented portfolios of employees on their roles and responsibilities at organizations.

4.1 LibRA: LinkedIn based Role Ambiguity

Defining LibRA. Drawing upon the theoretical definition of role ambiguity given in Section 2, we operationalize LinkedIn based Role Ambiguity (`LibRA`) as the *quantified differences in the self-explained roles and responsibilities of the individual against that posted by the company for the same role in the organization.* For this, we first obtain the self-explained job summary from an individual's LinkedIn profile. Then, for each role, we obtain the company described job description by manually conducting search engine queries of the specific role and the company. These job descriptions are typically posted on job posting websites, such as *Glassdoor, LinkedIn, Indeed*, and the *Google job search portal* — where the Google job search portal collates both exact and nearest matching job descriptions from multiple websites, including company's own website, LinkedIn, Glassdoor, Indeed, etc, and sorts them in relevance to the search query. For instance, Fig. 2) shows an example LinkedIn role description and company-published role description for the same role of Software Development Engineer at the same location of the company. Two coauthors independently obtained the nearest matching job description per role and per company — there were very few (<20) instances when

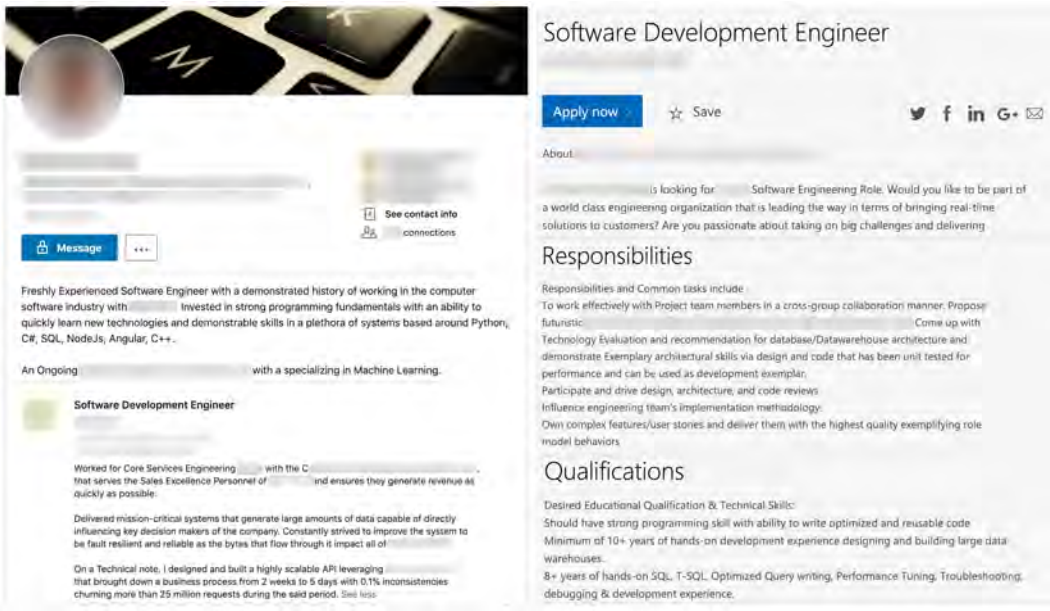


Fig. 2. For the same role (Software Development Engineer): (left) Role summary of an individual as described on LinkedIn, (right) Job description as posted on the company webpage

the two coauthors obtained different job descriptions, and when they did, the descriptions were very similar in the two websites, and the more relevant one was chosen.

Assessing LibRA. Next, towards computing LibRA, we first represent the above descriptions of self-reported LinkedIn job descriptions and the company described job descriptions in a multi-dimensional space of job aspects, for which we leverage O*NET. O*NET[3] is an online database and job ontology that contains a comprehensive list of jobs and their descriptions, elaborating on eight notable aspects of job role — *abilities, interests, knowledge, skills, work activities, work context, work styles, work values* (see Table 1 for brief descriptions). These aspects are grounded in literature and have been used in prior work to study employee behavior [141]. For every individual's role, we obtain their closest matching O*NET roles. For this, we adopt a semi-automatic approach of edit-distance based match, followed by manual evaluation and curation by two co-authors; the coauthors are familiar with, and are users of LinkedIn. For example, the closest match of *Software Development Engineer* is *Software Developers*.

Then, drawing on natural language analysis methods, we use word-embeddings, particularly pre-trained GloVe vectors [113, 123] to project the role descriptions of individuals and companies in a 50-dimensional word-vector space, so as to obtain rich lexico-semantic context surrounding the hand-curated job descriptors above [122]. We use cosine similarities to obtain two vector projections in the eight-dimensional job aspect space per individual i— 1) one that is obtained from their LinkedIn summary (v_1^i) and 2) one that is obtained from the same role's company description (v_2^i). Then, the overall LibRA is measured as the euclidean distance between v_1^i and v_2^i. To obtain the aspect-wise LibRA of an individual as the absolute difference per dimension of v_1^i and v_2^i. For instance, Fig. 3 show heatmaps of multi-dimensional role ambiguity of randomly selected 50 individuals

[3]O*Net (*onetonline.org*) is developed under the sponsorship of the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA).

Table 1. Job aspect types with their descriptions as obtained from O*Net.

Job Aspect	Description
Abilities	Enduring attributes of the individual that influence performance.
Interests	Preferences for work environments and outcomes.
Knowledge	Organized sets of principles and facts applying in general domains.
Skills	Developed capacities that facilitate learning or the more rapid acquisition of knowledge.
Work Activities	General types of job behaviors occurring on multiples jobs.
Work Context	Physical and social factors that influence the nature of work.
Work Styles	Personal characteristics that can affect how well someone performs a job.
Work Values	Global aspects of work that are independent to a person's satisfaction.

Fig. 3. Aspect-wise LibRA for a random set of 50 participants in two companies C_1 (above) and C_2 (below). These visualizations are an example comparison of LibRA within- and across- company employees

from two companies, C_1 and C_2 in our dataset. We find that some individuals show high or low role ambiguity across the aspects, but most individuals show high role ambiguity in one or more dimensions. While exploring the differences across multi-dimensional role ambiguity constructs remain a future research goal, we envision that such multi-dimensional role ambiguity [134] can benefit various stakeholders (employers or employees) through guided intervention to minimize their role ambiguity. This kind of interface is additionally inspired from prior HCI work aimed at facilitating employee satisfaction [33, 131].

4.2 Evaluating the Validity of LibRA Against Gold Standard

After defining and proposing a method to measure LibRA using LinkedIn data of individuals, we examine the validity of the measure. That is, we examine if the LibRA measure gets at least close to what the Role theory identifies as "role ambiguity". For this, drawing on modern validity theory [29], we compare the LibRA of the individuals against a gold standard validated survey on measuring role ambiguity. The Michigan Assessment of Organization survey instrument measures an individual's role ambiguity, role conflict, and role overload [107]. Corresponding to role ambiguity, the scale asks the participants to rate the four statements, "Most of the times I know what I have to do on my job", "On my job I know exactly what is expected of me", "I can usually predict what others will expect of me on my job", and "Most of the time, people make it clear what others expect of me", a 7-point Likert scale ranging from "Strongly Agree" to "Strongly Disagree".

We randomly sample a subset of 77 participants from our entire participant pool to answer the Michigan Assessment of Organization survey [107]. Correlating the survey-based role ambiguity with LibRA, we find Spearman's[4] correlation coefficient to be 0.22 ($p < 0.05$).

[4]Because the survey instrument on role ambiguity and our measure of LibRA measure role ambiguity in different scale and order, it makes sense to correlate the ranked (or relative) values rather than the raw values

Table 2. Summary of covariates used in the regression models.

Covariates	Value Type	Values / Distribution
Demographic Characteristics		
Gender	Categorical	Male | Female
Age	Continuous	Range (22:63), Mean = 35.24, Std. = 9.46
Education Level	Ordinal	4 values [College, Grad., Master's, Doctoral]
Job-Related Characteristics		
Income	Ordinal	7 values [<$25K, $25-50K, ... , >150K]
Tenure	Ordinal	10 values [<1 Y, 1Y, 2Y, ... 8Y, >8Y]
Supervisory Role	Boolean	Supervisor | Non-Supervisor
Job Type	Boolean	IT | Non-IT
Executive Function (Shipley scale)		
Fluid (Abstraction)	Continuous	Range (5:23), Mean = 16.91, Std. = 2.78
Crystallized (Vocabulary)	Continuous	Range (0.0:40.0), Mean = 33.38, Std. = 4.18
Personality Trait (BFI scale)		
Openness	Continuous	Range (1.7:5.0), Mean = 3.86, Std. = 0.54
Conscientiousness	Continuous	Range (1.7:5.0), Mean = 3.87, Std. = 0.66
Extraversion	Continuous	Range (1.7:5.0), Mean = 3.41, Std. = 0.67
Agreeableness	Continuous	Range (2.1:5.0), Mean = 3.85, Std. = 0.56
Neuroticism	Continuous	Range (1.1:4.7), Mean = 2.51, Std. = 0.77

Consequently, a statistically significant correlation does imply criterion validity, and hints at construct validity in our claim that LibRA does contain information that is also captured by gold standard, validated survey instruments on role ambiguity. However, we also acknowledge that the magnitude of correlation is moderate, which could be attributed to the differences in the measures (one is "perceived", and other is objectively measured) — we revisit some of the other nuances and factors that may cause these differences again later (Section 6 and 7) in this paper.

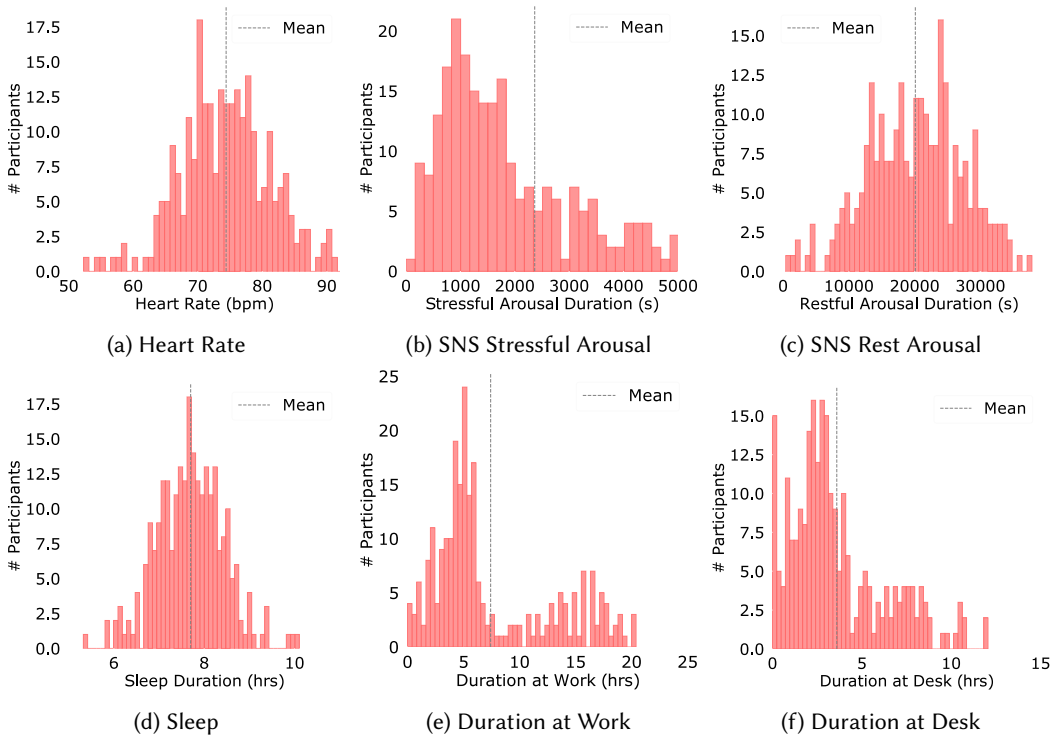
5 EXAMINING RELATIONSHIP OF LibRA WITH WELLBEING AND PERFORMANCE

This section revisits the hypotheses as outlined in section 2 towards establishing convergent validity of LibRA. For this, we study the relationship (and association) of LibRA with the passively sensed wellbeing measures, and the validated survey-based job proficiency measures. Because we are interested in studying the relationship, we consider linear regression models, which are known to provide easily interpretable associations in cases of conditionally monotone relationships with the outcome variable [32]. For every wellbeing or performance measure $\mathcal{M}$, we build linear regression models with $\mathcal{M}$ as the dependent variable, and LibRA as an independent variable, controlled for demographic, personality, and executive function measures per individual (see Equation 1). Our choice and inclusion of these covariates are motivated from prior literature [7, 15, 159]. Table 2 summarizes these covariates in their kind, and the values attained. For all the regression models, we use variance inflation factor (VIF) to eliminate multicollinearity of covariates (if any) [109]. For the ease of comparing the relative importance of the predictive variables in the regression models, we standardize their such that the variables have a mean of zero and standard deviation of one.

$$\mathcal{M} \sim gender + age + education\_level + income + supervisory\_role + tenure + job\_type + executive\_function + personality\_trait + LibRA \quad (1)$$

5.1 LibRA and Wellbeing

5.1.1 H_1: Greater role ambiguity is associated with greater heart rate. High heart rate is associated with an increase in stress [9, 64]. Caplan and Jones found that greater role ambiguity is associated with increased heart rate, which is identified as a major predictor of coronary heart rate [9, 21].



(a) Heart Rate (b) SNS Stressful Arousal (c) SNS Rest Arousal

(d) Sleep (e) Duration at Work (f) Duration at Desk

Fig. 4. Distribution of wellbeing measures as inferred via passive sensors.

We obtain the heart rate measures of the participants through the wearable sensor (see Section 3, Fig. 4 (a)). We fit a linear regression model with the average heart rate (HR) in the study period per individual. Given that exercise and physical activity has an association with heart rate [55], in addition to the covariates listed in Table 2, we control for the physical activity per participant. The regression model reveals a positive standardized coefficient (0.10) with statistical significance for LibRA (Table 3, Fig. 5 (a)). This observation supports our Hypothesis H_1.

5.1.2 H_2: Greater role ambiguity is associated with greater stressful arousal. Arousal is a physiological response that is related to one's heart rate variability, and is associated with stress, fatigue, and anxiety [35, 64]. These wellbeing measures are known to exacerbate in the presence of role ambiguity [1, 21]. In our project, the wearable sensor allows us to obtain participant arousal, particularly their Sympathetic Nervous System (SNS) arousal measures in a continuous fashion. In particular, for every individual, it scores the arousal level from restful to stressful on a scale of 1-100 at every three-minute granularity (Fig. 4 (b&c)). Here, the restful duration is when an individual relaxes or recovers from stress [3]. We build two separate regression models, one with median duration of high stressful arousal (75-100), and one with median duration of restful arousal (1-25) per individual. We find that LibRA shows a positive standardized coefficient (0.42) in the former model, and a negative standardized coefficient (-0.22) in the latter model (Table 3, Fig. 5 (b&c)). This suggests that individuals with high LibRA are more likely to show higher stressful arousal, and lower restful arousal. Therefore, our observations support H_2.

5.1.3 H_3: Greater role ambiguity is associated with decreased sleep. Sleep is an important attribute in an individual's wellbeing, and it reduces the negative impact of stress as well as improving

Table 3. Summary of standardized coefficients of regression models of wellbeing.

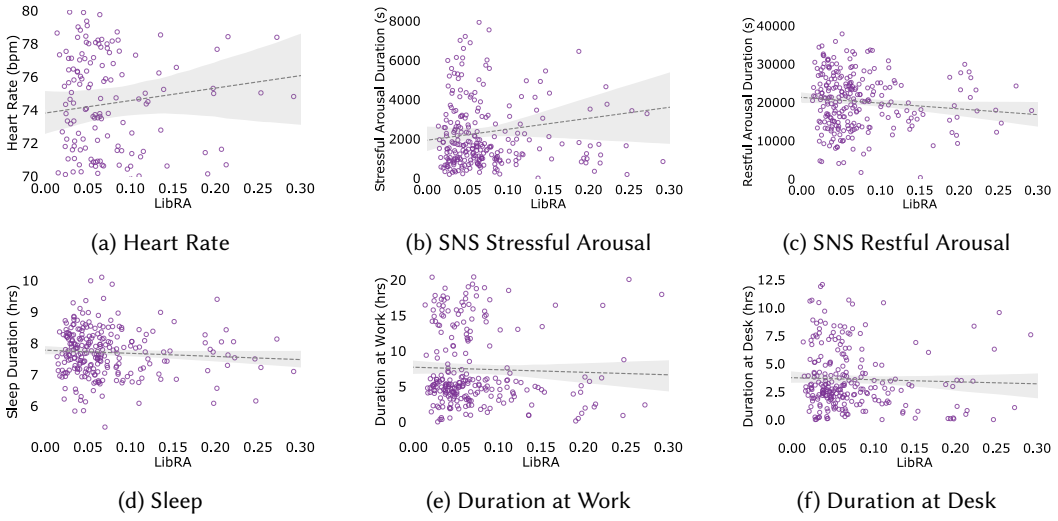
Covariates	Std. Coeff.	Covariates	Std. Coeff.	Covariates	Std. Coeff.
H_1 (Heart Rate)		H_2 (Arousal)			
$\mathcal{M}$ = Heart Rate, $R^2 = 0.16^*$		$\mathcal{M}$ = Stressful Duration, $R^2 = 0.65^*$		$\mathcal{M}$ = Restful Duration, $R^2 = 0.47^{**}$	
Exercise Duration	0.53^{**}	Age	0.69^{**}	Job: Non-IT	0.31^*
Shipley: Abs.	-0.81^*	Edu: Grad. School	-0.24^*	**LibRA**	-0.22^{***}
Agreeableness	0.91^*	Tenure: 4	-1.59^*		
Conscientiousness	-0.78^*	**LibRA**	0.42^{***}		
LibRA	0.10^*				
H_3 (Sleep)		H_4 (Work-Hours)			
$\mathcal{M}$ = Sleep Duration, $R^2 = 0.19^{***}$		$\mathcal{M}$ = Duration at Work		$\mathcal{M}$ = Duration at Desk	
Income: $50K-75K	0.21^*	Edu.: College	0.23^{***}	Duration at Work	0.18^*
Agreeableness	-0.14^*	Edu.: Grad. School	0.21^{***}	Edu: College	-0.09
Tenure: 7 Yrs.	-1.74^*	Income: $50K-75K	0.14^{***}	Edu: Grad.	-0.04
Job: Non-IT	0.15^{**}	Income: $100K-125K	-0.18^{***}	Edu: Master's	0.04
LibRA	-0.16^{***}	Shipley: Abs.	0.01^{***}	Income: $100K-125K	0.09^*
		Extraversion	0.09^{***}	Income: $125K-150K	0.08^*
		Conscientiousness	0.05^{***}	Tenure: <1 Yr.	-0.18^{***}
		Neuroticism	0.12^{**}	Tenure: 2 Yrs.	0.18^{***}
		Tenure: 6 Yrs.	-0.16^{***}	Tenure: 3 Yrs.	0.26^{***}
		Tenure: 7 Yrs.	-0.15^{***}	Tenure: 4 Yrs.	0.09^{***}
		Tenure: 8 Yrs.	-0.31^{***}	Tenure: 8 Yrs.	0.15^{***}
		Job: Non-IT	0.20^{***}	Job: Non-IT	-0.03^*
		LibRA	-0.41^{***}	**LibRA**	-0.12^{**}

overall health [13]. Given that stress reduces sleep, and sleep reduces stress, a stressed person is likely to sleep less [152]. If role ambiguity is stressful, we hypothesize that high role ambiguity will correspond with reduced sleep duration. The wearable sensor allows us to obtain participant sleep durations (see Fig. 4 (d)). We build a linear regression model with median duration of sleep per individual. We find that LibRA shows a negative standardized coefficient (-0.16) with statistical significance (Table 3, Fig. 5 (d)). Therefore, H_3 is supported in our dataset.

5.1.4 H_4: Greater role ambiguity is associated with decreased work hours. Role ambiguity is known to affect an individual's workplace behavior [112]. The bluetooth beacons sense if a participant is at work, at home, or commute, and within work; it additionally captures the duration the participant is at- and away from- desk. We build two regression models, one with the duration at work, and one with the duration at desk, when at work (this model additionally controlled for duration at work). Here, we find that both of these dependent variables show heavy-tailed distributions (see Fig. 4 (e&f). For both of these distributions, Chi-squared tests could not reject the null hypotheses that they were significantly different from a Poisson distribution ($p > 0.05$). Therefore, instead of using purely linear regression models, we build negative binomial regression models [67], ones that essentially regress the logarithm of the dependent variables with the independent variables [67]. We prefer negative binomial regression over poisson regression because we find the presence of over-dispersion in the distribution of both duration at work and duration at desk (Fig. 4 (e&f)) [30]). We find that LibRA shows a negative standardized coefficient in both the models (-0.41 for duration at work, and -0.12 for duration at desk, Table 3, Fig. 5 (e&f)). This suggests that individuals with high LibRA are not only less likely to spend time at work, but also less likely to spend time at desk when at work. These observations support H_4.

5.2 LibRA and Job Performance

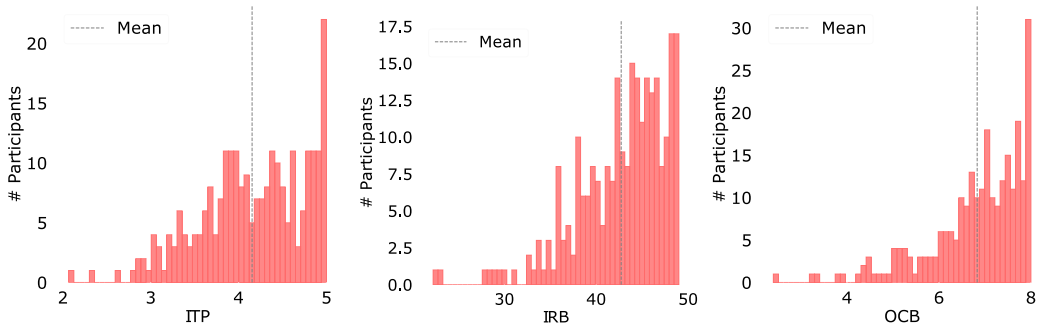
5.2.1 H_5: Greater role ambiguity is associated with lower task performance. We administered two survey scales of In-Role Behavior (IRB) and Individual Task Performance (ITP) three times a week,



(a) Heart Rate (b) SNS Stressful Arousal (c) SNS Restful Arousal

(d) Sleep (e) Duration at Work (f) Duration at Desk

Fig. 5. Scatter plots of demonstrating the distribution of wellbeing attributes against LibRA. LibRA is positively associated with heart rate, stressful arousal, and negatively associated with restful arousal, sleep, duration at work, and duration at desk. In sum, increase in LibRA is associated with depleted wellbeing.



(a) ITP (b) IRB (c) OCB

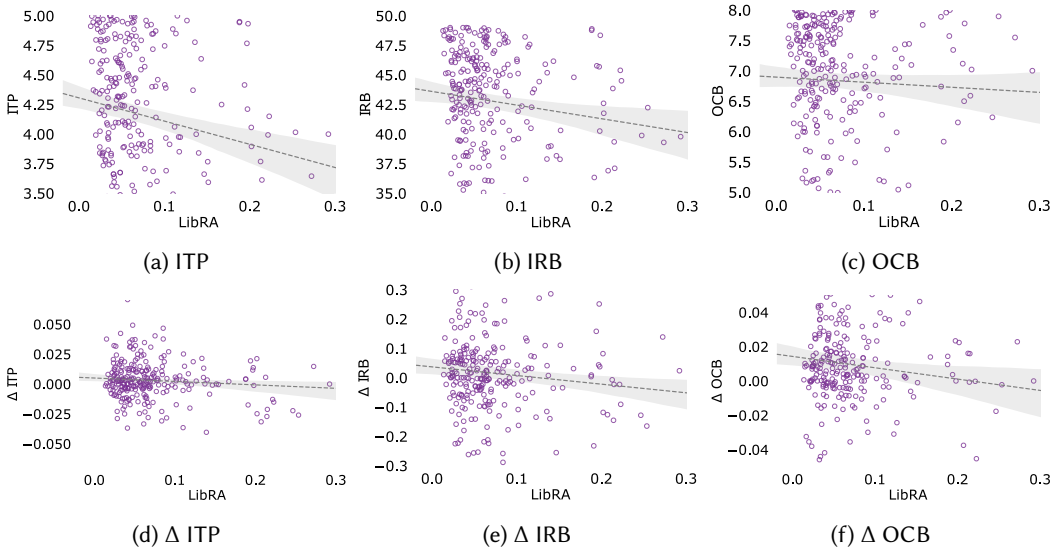
Fig. 6. Distribution of Performance measures via job performance surveys.

to periodically obtain the self-assessed task performance of the participants (see Section 3, Fig. 6 (a&b)). For both these measures, we build two linear regression models each — one that uses an aggregated (median) value of task performance, and one that uses a change in task performance over the duration of the study. We find that LibRA shows a negative association with both *aggregated ITP* (-0.33) and *change in ITP* (-0.20) per individual. Similarly, LibRA also shows a negative association with both *aggregated IRB* (-0.29) and *change in IRB* (-0.20) per individual (Table 4, Fig.7 (a&b, d&e)). Together, these observations suggest that individuals with higher LibRA not only have a greater likelihood of performing badly at work, but also their performance worsens over time. Therefore, our observations support H_5.

5.2.2 H_6: Greater role ambiguity is associated with lower organizational citizenship behavior. We administered the Organizational Citizenship Behavior (OCB) scale three times a week, to periodically obtain the self-assessed organizational citizenship behavior of the participants (Fig. 6 (c)). Like the

Table 4. Summary of standardized coefficients of regression models of task performance.

Covariates	Std. Coeff.	Covariates	Std. Coeff.	Covariates	Std. Coeff.
H_5 (Task Performance)				H_6 (Org. Citizenship Behavior)	
$\mathcal{M}$ = ITP, R^2 = 0.29***		$\mathcal{M}$ = IRB, R^2 = 0.29***		$\mathcal{M}$ = OCB, R^2 = 0.24***	
Income: L	-0.38*	Openness	0.13**	Supervisor: Yes	0.24***
Income: Q	0.40**	Consc.	1.13*	Extraversion	0.34***
Openness	1.07*	Tenure: 8	0.17*	Tenure: 6	-0.14*
Consc.	1.30***	**LibRA**	-0.29*	Tenure: 7	-0.20**
Tenure: 6	-0.15*			**LibRA**	-0.10**
LibRA	-0.33***				
$\mathcal{M}$ = Δ ITP, R^2 = 0.13*		$\mathcal{M}$ = Δ IRB, R^2 = 0.17***		$\mathcal{M}$ = Δ OCB, R^2 = 0.22*	
Extraversion	0.69*	Openness	0.91**	Supervisor: Yes	-0.26*
Consc.	-1.37***	Consc.	-0.84*	Agreeableness	-1.80*
LibRA	-0.20*	Tenure: 7	-0.19*	Tenure: 5	0.21*
		Tenure: 8	-0.26**	**LibRA**	-0.25***
		Tenure: 9	-0.18**		
		LibRA	-0.20**		



(a) ITP (b) IRB (c) OCB

(d) Δ ITP (e) Δ IRB (f) Δ OCB

Fig. 7. Scatter plots of demonstrating the distribution of job performance measures against LibRA. LibRA is negatively associated with ITP, IRB, OCB, Δ ITP, Δ IRB, and Δ OCB. In sum, an increase in LibRA is associated with both decreased job performance as well as reduced job performance over time.

above, we build two linear regression models — one that uses an aggregated (median) value of OCB, and one that uses a change in OCB over the duration of the study. We find that LibRA shows a negative association with both aggregated OCB and change in OCB per individual (Table 4, Fig.7 (c&f)). These observations suggest that individuals with higher LibRA show a greater likelihood of poorer OCB, which also worsens over time — a tendency associated with being disinclined to be altruistic or help colleagues at workplace. Therefore, our observations support H_6.

6 INVESTIGATING THE FACTORS AFFECTING LibRA

This final section studies the factors that contribute to the LinkedIn based role ambiguity (LibRA) assessment. Specifically, we investigate the extent to which appropriating data shared online (on

a professional social networking service, LinkedIn) may bring forth new dimensions to consider while employing LibRA for practical use, and what might contribute to observed differences in LibRA. To do this, we draw from various literature to situate our observations.

First, we seek to quantitatively study the relationship of LibRA with observable and intrinsic attributes of an individual. Using the same covariates as in Table 2, we fit one's LibRA as the dependent variable in a series of statistical models. Our rationale to study this rests on prior work that found demographic and intrinsic traits affecting self-disclosure behavior on LinkedIn, which may lead to differences in LibRA [63, 144, 146, 150]. We build multiple regression models (both linear and non-linear), but find no significance in the relationship ($p > 0.1$) in either the regression fit or the variable coefficients. ANOVA F-test per covariate and LibRA reinforced our confidence in this finding that there is no significant relationship in the variability of observable traits influencing LibRA. This aligns with previous literature that claims role ambiguity is independent of individual traits, rather than an outcome of a number of factors such as mentor-mentee relationship, working alliance, organizational structure, and organizational communication [87]. Nevertheless, because LibRA is inferred from social media data, specifically LinkedIn, we recognize that numerous mediators can confound the self-presentation behavior of an individual on LinkedIn (even after controlling for their intrinsic demographic and personality traits). We delve deeper into this consideration based on a qualitative examination of a sample of our dataset as described below.

We intend to compare and study the self-presentation behavior, accounting for the between-individual differences in self-reported and assessed traits of demographic, personality, executive function, and work role-related characteristics. Therefore, with these characteristics as covariates (see Table 2), we draw on matching techniques from causal inference [71, 124] to match individuals using Mahalanobis Distance Matching [59]. We separately match pairs of individuals who belong to IT roles, and who belong to non-IT roles. Fig. 8 plots the pair-wise Mahalanobis distances and the absolute differences in their role ambiguities. We focus on those individuals (shaded region in Fig. 8) who are similar in their individual attributes but show high differences in their LibRA – we sample the top 10th percentile of pairs of individuals in IT and non-IT each.

Next, among the individuals in the above sample, we manually look at their (public) LinkedIn job and profile descriptions. While these individuals are very similar in their personality, demographic traits, and their role in the company (because of matching), in terms of their self-presentation behavior on LinkedIn, we find differences in their style of writing (also highlighted in the Fig. 8 examples). For example, one writes an extremely short description compared to their matched other, who writes a longer description with much more detail. Another example includes only technical-skills or the tasks that they are assigned at work (e.g., *Java, business development*), compared to their matched other, who additionally describes their non-technical and people skills and abilities (e.g., *accomplished, dynamic*). Given the affordances and the uniqueness of LinkedIn as a professional social networking platform, we deduce a few plausible reasons that can potentially influence the virtual self-presentation of the individuals, and in turn, lead to varied inferred role ambiguity. We discuss these factors, which are not disjointed and could be inter-related:

Individuals' Organizational Behavior. Individuals who are looking for newer jobs or endeavors possibly write a more detailed portfolio on their LinkedIn profiles, whereas individuals who are generally "settled" are not as active in providing detailed descriptions [135]. This could also be a *different type* of job than what they are currently involved at altogether as well. In other words, the settled people may have different jobs currently compared to they were hired, e.g., through promotion or lateral moves within the company. An alternative conjecture could also be that, only a few individuals write and "highlight" their *work experience*, rather than describing their responsibilities and tasks at the workplace, for example, "I have 25 years of Health Care Provider experience in revenue cycle selling and managing outsourced health care accounts, receivable solutions [..]". We find individuals who describe their role with people skills and proficiencies beyond their tasks, such as "I can effectively cope with change, shift gears comfortably, and bring

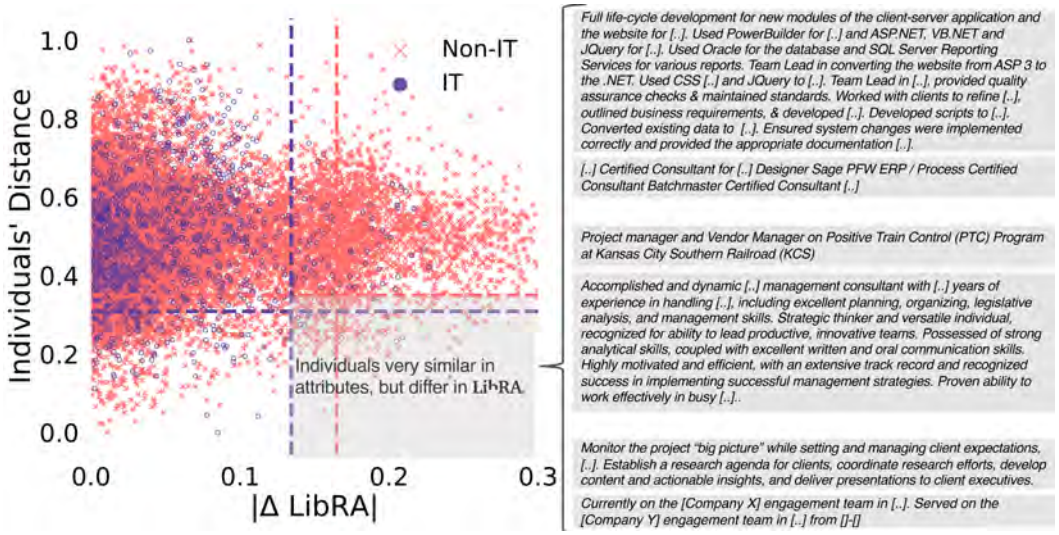


Fig. 8. Pair-wise differences in individual attributes and corresponding differences in LibRA. Example excerpts show the differences in LinkedIn descriptions of pairs of individuals with very similar individual attributes (low differences), but large differences in LibRA.

a point of view to the leadership", and those who describe their attitude towards initiativeness, "I am always willing to help especially if there is a problem to be solved, and my behavior is a mix of light-heartedness and a drive to put into practice everything I have learned.". These could be individuals who exhibit proactive behaviors in the organizations [27]: they show anticipatory, change-oriented and self-initiated behavior in situations and tend to act in advance of a future situation, rather than just react later. This may also indicate that although these individuals have high role ambiguity, they show desirable individual characteristics (proactive behavior and leadership traits) in organizations [8, 27]. Exploring these aspects further is of future research interest.

Individual-Intrinsic Factors. Prior research has observed that many individuals tend to self-promote and appear honest and less deceptive on their professional social networking profiles [60, 151]. However, the degree and the way in which they self-present themselves can vary. Given the context of professional choices and career development, we can look at it from the perspective of growth versus fixed mindset [39]. Those with "fixed mindset" believe their abilities are innate, whereas the ones with "growth mindset" believe that abilities can be acquired via investing effort and study. For instance, an individual describes themself as, "a motivated and hardworking professional looking to improve my skills and abilities." Although mindset and personality traits are somewhat related, mindset can reshape over time and through interactions [2]. Complementary research directions have also coined "benefit mindset", and "global mindset", "productive mindset", and "defensive mindset", all illustrating a variety of intrinsic behaviors of individuals that contribute to their skill development, proficiency, and self-presentation in organizations [18, 62]. We conjecture that similar traits permeate into their online self-promotion practices on LinkedIn.

Job-Related Factors. Literature has demonstrated the importance of job titles in organizations [147]. LibRA assessments of an individual are derived from the job titles of the individuals. However, if the job titles themselves are ambiguous then that inherently adds ambiguity to the role of the individual. In fact, we find pairs of individuals where one is an "Associate", while the other is a "Specialist" — both of these titles are pretty generic, and do not convey much information to the employees. In contrast, the fact that recently companies are coming up with "cool" job titles (e.g., *ninja*) to

gain visibility and distinctiveness can add other complexities to role ambiguity [126]. As Utz and Breuer recently noted that one's career orientation, type of role or organizational sector, influence their behavior or use of LinkedIn— for example some sectors may require more referrals or information than others, thereby implicitly demanding greater activity from the individuals [149, 164]. Additionally, some individuals may be working on confidential projects and they are bound by nondisclosure agreements. Further, the role in a company and size of a company can influence the self-description behavior of individuals [164]. That is, even with common and similar job titles, individuals at large enterprises may not feel the need to describe their role in as much as detail as those at startups and mid-size organizations [95]. Compounding this difference in company size, some companies may encourage the use of LinkedIn among employees to improve the image of the company, or may even render the platform as a mandatory in-company communication tool, thereby influencing the LinkedIn use behavior of their employees [151].

Audience, Privacy, and Platform Factors. Finally, the familiarity or the use of LinkedIn as a platform may vary across individuals. Two participants in our sample described what their company does, rather than what their role is, such as, "[Company] specializes in [..] and works with companies that offer [..] service. [Company] has over 40 years of experience in the industry and operates groups of 10 to 1000 people [..]." In addition, LinkedIn is a professional social networking platform that also functions as a marketplace for job seekers. Individuals tend to share credible information because they have a conceptualization of an "invisible audience" [10], and since LinkedIn is a public space, they do not want to appear as dishonest [60]. At the same time, as discussed in Ghoshray's work, employee surveillance and employee's subjective expectation of privacy shares a competing relationship, and the sheer perception of being "surveilled" can influence one's self-disclosure behavior on the platform [51, 73, 146]. Further, employee's own mental models about LinkedIn privacy might be a factor behind what they share [22].

In summary, LibRA is based on self-presentation on a professional social media site, Linkedin. As such, it is subject to variability in self-presentation and motivation found in the population, such as differences in organizational behavior, differences in job status (e.g. looking for a new job vs remaining established), differences in values (e.g. "fixed" vs "growth" mindset), differences in the context of the job (e.g. a software engineer at a small firm vs a large firm) and the assigned job title, and differences in how individuals perceive the positives and drawbacks of their professional information in a publicly accessible space. These differences should be considered when applying LibRA in assessing role ambiguity. We discuss the consequences and implications associated with such a measure in the next section.

7 DISCUSSION

7.1 Theoretical Implications

This work measures role ambiguity (LibRA) for information workers with a diverse set of individual difference attributes using their self-described portfolios as shared on professional social networking website (LinkedIn). Traditionally, registries and census organizations have served as analogous source of data for people's professional portfolios. Our work reveals the feasibility of measuring a role related construct (here LibRA) at scale via a previously unexplored, low-cost, and unobtrusive source of data. Management and economics research is advancing in ways that can use this data to operationalize and derive existing measures (e.g., role constructs and role stressors) in novel ways. Thereby, our work revisits old questions in labor economics. While existing efforts so far have been limited to utilizing statistical numbers such as salary distribution, unemployment rates, and so on. Our work can potentially complement these with richer information on employee job satisfaction at scale.

This work lays the foundation of studying employee organizational psychology and behavior through unobtrusive online data sources, that set up marketplaces for employees, such as other

professional networking websites such as Meetup, Xing, Jobcase, etc. Because our method is platform-agnostic, and such career portfolio like in LinkedIn is universally available, our work can be easy to replicate in other contexts such as diverse workplaces and organizations, and other types of situated communities. In addition, our work combines organizational psychology, organizational behavior, and organizational strategy streams of literature, and can further be used to advance our understanding of coping mechanisms, incentives, and job satisfaction in general at workplaces, by adopting a technology-focused and technology-driven lens.

Because this work uses individuals' self-described portfolios of job roles and responsibilities, we can to objectively assess the differences in "what the individual considers and self-describes themselves to be doing", and "what the company hired them for, or what their job description states". That is, the individual may only be showing normative and socially influenced behavior at their work, or show there is information gap, or demonstrate they intend to invest more effort to learn and gather experience themselves. These are behaviors that are detectable oblivious to the presence of role ambiguity. Such "unaware role ambiguities", is challenging to capture via survey instruments as they are tuned to measure the "perceived role ambiguity". Language can reflect differences in personal traits as well as situational ones [53]. This additionally makes our measure less subjectively biased than traditional methods of measuring role ambiguity.

7.2 Practical and Technological Implications

7.2.1 Individual-Centric Implications. From a practical implication perspective for an *individual*, our work can be used to develop self-reflection tools for employees to mitigate their intrinsic bias in perceived role ambiguity. This can help them continually assess themselves on their skillset and productivity at work. Such self-reflection tools can include within and across organizational role comparisons, and also within and across industrial role comparisons. These can potentially benefit the employees to have more streamlined information on their end that reduces their job search costs and effort, and enhances their wellbeing. In addition, describing tasks or job role in itself is partially a self-reflection process, and a tool that scores for a type of description will help minimize bias, and also help the employees to identify sources of their role ambiguity.

Further, self-reflection is known to have a positive influence on productivity and job satisfaction [34]. Integrating self-reflection tools with our approach would facilitate automated (self)-assessment of one's skillset, interest, and adaptability to an organizational role, and indirectly help them estimate their productivity, wellbeing, and job satisfaction at both their current as well as a future potential workplace. By logging roles, responsibilities, and tasks in a longitudinal fashion, an individual can assess their professional growth and development, and can also be prompted with recommendations for skill training wherever necessary. For the individuals who want to seek professional career-related advice, these logs can function as a diary-style data source to professional mentors and career counselors for better understanding of one's career trajectory, beyond the information presented in a resume.

7.2.2 Organization-Centric Implications. Presently job and skillset training at organizations is not streamlined [108]. Either they train a lot of employees in a batch, or they mentor them individually. However, with more information regarding how employees perceive their role, employers can identify the area of training required that will reduce role ambiguity and enhance the productivity of employees. With our method, since the time to identify such role ambiguity gaps can be reduced, both training costs and employee wellbeing costs for the organizations can be reduced. This in turn, can improve employee retention for companies by identifying turnover intentions.

Aligning with and confirming the literature [87], our findings suggest that LinkedIn inferred role ambiguity (`LibRA`) is not dependent on individual differences such as personality, gender, supervisory role, and executive function. This can inform organizations how these roles or titles can be transformed to match skill-level, task-assignment level, and incentive-level restructuring. In addition, this work calls for more careful development of job descriptions. Organizations can involve

team- or sector- level staff in curating of job descriptions that are more attuned to the responsibilities and skills of the employees, and can dynamically update the descriptions in accordance with the necessity [69, 83]. Together, these measures can help improve job attractiveness and employee satisfaction in the company.

The interest in human resource management is still nascent but promising in the HCI and design community. In fact, cross-disciplinary literature pertaining to workplaces and online technologies provide potential use-cases urging the attention of designers [132]. Our work has implications towards designing and developing organization-centric technologies:

(1) First, tools can be built that suggest carefully chosen, fine-tuned job titles to companies, based on our measure of `LibRA` [6, 56]. This is particularly important because younger organizations sometimes offer (higher ranking or impressive-sounding) titles to employees in lieu of higher salaries, but this strategy has been reported to backfire due to increased role ambiguity and in turn, affecting employee productivity and wellbeing [126]. Adoption of tools that inform organizations about existing ambiguities in specific job roles, therefore, has the potential to make the workplace and individual roles more conducive to effective coping against workplace stress [88]. Moreover, professional social networking platforms (such as LinkedIn) are already heavily used to recruit by job agencies and resume matching consultancies [84]. Such agencies can leverage the insights gained from our approach to both match as well as recommend suitable jobs to prospective employees that are likely to result in lower role ambiguity.

(2) Second, our work can help design workplace tools and dashboards to enhance organization "health" or functioning. Such dashboards can unobtrusively and proactively assess employee role ambiguities at scale, taking employees' privacy considerations into account. In fact, many companies already provide their employees with internal social media platforms [36], online engagement forums, or even email profile description spaces, where they can regularly update their self-explained expertise and role descriptions, along with manager or peer-appraised testimonials. By leveraging such internal datasets, management in companies can potentially adopt these dashboards to gauge role ambiguity to make informed role matching for teams that require internal hires for open positions or internal role/team swappings. Companies can also restructure and reassign current employees with appropriate incentivization and compensation on their task and workload.

It is also important to note here that our results showed that *role ambiguity may not necessarily be "bad"*. It is possible that individuals who demonstrate desirable organizational characteristics, such as proactivity and initiative [27], may show high role ambiguity, simply because of their desire to self-present in a particular way. Therefore, we suggest caution in how `LibRA` is made actionable by companies, especially in the light of the many possibilities to build the above organization-centric technologies. We suggest that companies should not only encourage and provide rewards for these type of employees because they bring role and skill diversity to the organizational culture, but also consider shepherding these individuals with better coping strategies so that they deal better with their wellbeing concerns that are attributed to an underlying role ambiguity [86].

7.3 Social Computing Implications

Our work also has implications for *social computing system* design. Platforms such as LinkedIn cater to both individuals by recommending them jobs, as well as to companies by recommending them individuals. Their recommendation algorithms can leverage the quantified measure of role ambiguity (`LibRA`), complementing and going beyond general skills and experience matching. In addition, social computing platforms can aggregate role ambiguities between organizations, and within organizations across teams. This will add more transparency and objective information on company experiences, complementing the review websites (such as Glassdoor), which tend to be heavily polarized or biased on negative experiences [82]. These data sources will benefit both the job seekers as well as the employers in evaluating, understanding, and implementing measures to improve the work experience.

Finally, LinkedIn already enables individuals to gauge their "professional value" based on their profile stats [151]. An added feature to that could be a measure like LibRA, and guided recommendations on the basis of one's weaknesses (in terms of role ambiguity) to online training (such as Lynda[5]), or with classes at local third party training centers. For privacy-preserving purposes, LinkedIn anonymizes one's list of followers [151], but this also compounds the fact that there is no structured way to know "who sees what on Linkedin", adding complexity in terms of the audience is a problem that an individual faces [68, 91, 96]. However, the platform can adopt design changes such as allowing individuals with diversified interests to create multiple professional personas for different audiences. For example, someone who is both a Software Engineer as well as a part-time Physics Tutor, may self-promote their expertise and gain visibility in both the disciplines but to different controlled audiences [80], who can assess their role-related constructs only on the discipline that they are interested in.

7.4 Ethical, Privacy, Social, and Policy Implications

Back in 2014, when Zhang et al. studied "creepiness" and privacy concerns related to social media use by workplace professionals, they found concerns shifting from boundary regulation to behavior tracking by social media platforms for targeted advertising [163]. However, social media- and the web- based behavioral inferences has evolved tremendously since then, and also come under ethical and political scrutiny for privacy breaches such as the Cambridge Analytica scandal on Facebook [19]. Moreover, our work renews attention to the challenges that may arise when employee data is appropriated for workplace surveillance; as Van Dijck's research noted, "LinkedIn's functionality goes beyond its self-claimed ambition as a professional matchmaker, and ventures into behavioral monitoring." In fact, with research like ours, use of people's online self-presentation to infer their offline behavior (with high-risk decision outcomes such as one's profession or career) augments several complexities to one's perception of ethics and privacy, and consequently their behavior on social media. We discuss some of these challenges below.

Although our work leverages public social media of individuals, it raises new questions on the *privacy-breach of individual information*. An employee's motivations and expectations for LinkedIn might have been only to network or to browse jobs, and they may be well unaware that their published portfolio may also be used to analyze their present or future role-ambiguity and measures of organizational fit or job security [5, 31, 151, 163], which the individuals may not feel comfortable about, especially when this information is made accessible to their employers. Further, *this work is not intended to facilitate employer surveillance*, which shares a competing relationship with employee's subjective expectation of privacy [28, 51, 137].

More elaborately, per Goffman et al.'s theory of self-presentation, individuals may present two kinds of information — one that they intend to "give off", and one that "leaks through" without any intention [54, 102]. This implies that both of the perspectives may be present in our sort of research. Publishing role descriptions as online portfolios on public social media platforms like LinkedIn benefits the individuals in many ways, but research such as this may also abuse their data without their consent or awareness. For instance, employers may make inferences about role ambiguity and subsequent job satisfaction to make decisions on rewarding, promotions of employees, or even employee retention and layoffs. Therefore, to regulate such practices via the use of social media data, employee right protection agencies and lawmaking bodies such as the Department of Labor in the United States should consider making guidelines on how organizations engage in data-driven decision making regarding their workforce [73]. This work calls for constitutional jurisprudence in terms of employee social media rights and employer surveillance [51, 73].

Additionally, different companies have different kinds of expectations, history, culture, structure, and needs in their organizations that are latent and beyond what role descriptions say [43, 100, 132]. These factors, alongside platform-related and individuals' intrinsic factors that may impact their

[5] *linkedin.com/company/lynda-com/*, Accessed 2019-03-21

role ambiguity assessments, should be accounted for before making decisions merely on any sort of data-driven form of inferred role ambiguity.

On the contrary, *individuals may also start gaming the system*, and describe themselves in language that is more attuned with their role descriptions at work to gain professional advantages [151]. Such deceptive behavior calls for action for stakeholders with diverse interests ranging from academia and industry, as this adds complexities, and may even rigorously change the whole social computing ecosystem on LinkedIn compared to how it is used now. The platform may consider bringing moderation of content or user flair/karma (such as on Reddit) to enable that only credible information is shared on the platform. Presently, LinkedIn already includes features such as testimonials and recommendations that may be leveraged to counter such behavior on LinkedIn. However, such measures are biased as well and can cause Matthew Effect (*the rich get richer, and the poor get poorer*) [101], so accounting for them needs additional consideration.

In addition to the above, our work is only able to measure role ambiguities for those who are on LinkedIn, predominantly in white-collar jobs. According to Pew Survey Reports, 25% of U.S. adults are on LinkedIn, and the demographic is skewed towards the younger, urban, and college-graduate individuals [114]. It is likely that only "privileged" individuals can benefit from these kinds of online data- or technology- driven measures to advance and positively impact their job outcomes and wellbeing, e.g., via the self-reflection tools we discussed above. Consequently, those who are not on the platform (which could due to their socio-economic conditions, e.g., the vast majority of blue-collar job workers, or by choice), may feel compelled to use it owing to social and professional pressures of being on it. This adds to the complexities that prior work identified regarding *digital inequalities in job-seeking and job summarization behavior on the internet* [57, 76, 164].

7.5 Limitations and Future Directions

We recognize some limitations in this work. As mentioned, our approach of assessing role ambiguity applies to only those individuals who are on LinkedIn, and what individuals describe themselves on LinkedIn. Role ambiguity could be because of the role itself, or because of the inherent biases, intrinsic attributes, and online platform and self-presentation choices of the employees, as we noted in Section 6. Besides, the problem of honesty and deception on LinkedIn [60] remains underexplored, and should be accounted for when inferring data-driven workplace outcomes such as `LibRA`.

Our work is limited by the affordances and use-behavior of LinkedIn as a professional social networking platform. It is constrained by how updated the self-described portfolios of individuals are, and we only include a snapshot of LinkedIn profiles of individuals during the period of our study. However, an individual's role or perceived role in workplace is likely to change, and we do not account for any such internal changes or any internal communication that the management or the supervisors in the organizations may have made the employee aware about. Further, we note that causal inferences cannot be drawn based upon our study. Future work can involve experimental and quasi-experimental setting to study any sort of causal and temporal relationships of how role ambiguity, and LinkedIn profile update of individuals varies over time, and how does it affect their wellbeing, performance, and satisfaction in a workplace. Technically, our approach to quantify `LibRA` relies upon external sources of job description and manual inspection. Although such a process is challenging to be scaled, in practical scenarios such additional data can be obtained far more easily at low cost (both time and monetary) by the actual stakeholders and users of `LibRA` — individuals can use it against their own job descriptions, and organizations would have the repository of job descriptions to match against individuals, which are only periodically updated as new job titles are not created at rampant pace.

Despite using passively sensed wellbeing measures, this work suffers from the limitations as any other cross-sectional study that employs snapshots of data or self-reports at a point of time. Because we only study the relationship of `LibRA` with self-assessed job performance, we cannot make conclusive claims about its relationship with the employer or other stakeholders, such as

supervisors-, peers-, and subordinates- assessed performance or rewarding at the workplace. Future research can model their studies on the cognitive model of stress that centers around repeating the cyclic process of appraisal, coping, and reappraisal [145].

8 CONCLUSION

This paper proposed a methodology to quantitatively estimate role ambiguity via unobtrusively gathered data from LinkedIn. Our dataset consisted of consented LinkedIn data of 267 information workers in the U.S. who are participants in a multimodal sensing study of job proficiency. We computed LinkedIn based Role Ambiguity (LibRA) as a difference in one's self-described roles (on LinkedIn) and the company-published job description of the same role. In particular, we measured these differences using word-embeddings on the multiple dimensions of job aspects across *abilities*, *interests*, *knowledge*, *skills*, *work activities*, *work context*, *work styles*, and *work values*. Aligning with a set of theory-driven hypotheses, we find that greater LibRA is associated with depleted wellbeing, such as increased heart rate, increased arousal, decreased sleep, and higher stress. In addition, LibRA is associated with lower job performance such as decreased organizational citizenship behavior and decreased task performance. Finally, we explored the self-presentation behavior and social computing platform-specific nuances and factors that need to be accounted if measures like LibRA are to be used in practice. This work can help fill gaps in state-of-the-art assessments of role ambiguity, and we discussed the potential of this measure in building novel technology-mediated strategies to combat role ambiguity, and improve efficiency in organizations.

9 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 for their invaluable contributions in realizing the goals of this project. We also thank Sindhu Ernala, Vedant Das Swain, Shagun Jhaver, Dong Whi Yoo, Stevie Chancellor, and the members of the Social Dynamics and Wellbeing Lab at Georgia Tech for their feedback.

REFERENCES

[1] David J Abramis and Long Beach. 2017. Work Role Ambiguity , Job Satisfaction , and Job Performance : Meta-Analyses and Review. February (2017). https://doi.org/10.2466/pr0.1994.75.3f.1411
[2] Peter Aldhous and Carol Dweck. 2008. Free your mind and watch it grow.
[3] Garmin Health API. 2018. http://developer.garmin.com/health-api/overview/. Accessed: 2018-11-01.
[4] Manager REST API. 2018. Employer Access to Employee Social Media: Applicant Screening, "Friend" Requests and Workplace Investigations. Accessed: 2018-11-01.
[5] Anne Archambault and Jonathan Grudin. 2012. A longitudinal study of facebook, linkedin, & twitter use. In *Proc. CHI*.
[6] James N Baron and William T Bielby. 1986. The proliferation of job titles in organizations. *Administrative Science Quarterly* (1986), 561–586.
[7] Murray R Barrick and Michael K Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology* 44, 1 (1991), 1–26.
[8] Thomas S Bateman and J Michael Crant. 1993. The proactive component of organizational behavior: A measure and correlates. *Journal of organizational behavior* 14, 2 (1993), 103–118.
[9] Athanase Benetos, Annie Rudnichi, Frédérique Thomas, Michel Safar, and Louis Guize. 1999. Influence of heart rate on mortality in a French population: role of age, gender, and blood pressure. *Hypertension* 33, 1 (1999), 44–52.
[10] 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.
[11] Bruce J Biddle. 1986. Recent developments in role theory. *Annual review of sociology* 12, 1 (1986), 67–92.

[12] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D'Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. 2019. Prediction of Mood Instability with Passive Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019).

[13] Jessica M Blaxton, Cindy S Bergeman, Brenda R Whitehead, Marcia E Braun, and Jessic D Payne. 2017. Relationships among nightly sleep quality, daily stress, and daily affect. *The Journals of Gerontology: Series B* 72, 3 (2017), 363–372.

[14] Walter C Borman and SM Motowidlo. 1993. Expanding the criterion domain to include elements of contextual performance. *Personnel Selection in Organizations; San Francisco: Jossey-Bass* (1993), 71.

[15] Chieh-Chen Bowen, Janet K Swim, and Rick R Jacobs. 2000. Evaluating Gender Biases on Actual Job Performance of Real People: A Meta-Analysis 1. *Journal of Applied Social Psychology* 30, 10 (2000), 2194–2215.

[16] James A Breaugh and Joseph P Colihan. 1994. Measuring facets of job ambiguity: Construct validity evidence. *Journal of applied psychology* 79, 2 (1994), 191.

[17] Michael J Brzozowski. 2009. WaterCooler: exploring an organization through enterprise social media. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 219–228.

[18] Ashley Buchanan and Margaret L Kern. 2017. The benefit mindset: The psychology of contribution and everyday leadership. *International Journal of Wellbeing* 7, 1 (2017).

[19] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian* 17 (2018).

[20] Ralf Caers and Vanessa Castelyns. 2011. LinkedIn and Facebook in Belgium: The influences and biases of social network sites in recruitment and selection procedures. *Social Science Computer Review* 29, 4 (2011), 437–448.

[21] Robert D Caplan and Kenneth W Jones. 1975. Effects of work load, role ambiguity, and type A personality on anxiety, depression, and heart rate. *Journal of applied psychology* 60, 6 (1975), 713.

[22] João Caramujo and Alberto Manuel Rodrigues da Silva. 2015. Analyzing privacy policies based on a privacy-aware profile: The Facebook and LinkedIn case studies. In *2015 IEEE 17th Conference on Business Informatics*.

[23] Raymond Bernard Cattell. 1987. *Intelligence: Its structure, growth and action*. Vol. 35. Elsevier.

[24] Stevie Chancellor and Scott Counts. 2018. Measuring Employment Demand Using Internet Search Data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 122.

[25] Esther Chang and Karen Hancock. 2003. Role stress and role ambiguity in new nursing graduates in Australia. *Nursing & health sciences* 5, 2 (2003), 155–163.

[26] Jose M Cortina and Joseph N Luchman. 2012. Personnel selection and employee performance. *Handbook of Psychology, Second Edition* 12 (2012).

[27] J Michael Crant. 2000. Proactive behavior in organizations. *Journal of management* 26, 3 (2000), 435–462.

[28] Melissa Crespo and Christine E Lyon. 2015. Employer Access to Employee Social Media: Applicant Screening, "Friend" Requests and Workplace Investigations. *Compliance, Employment Law, Privacy* (2015).

[29] Linda Crocker and James Algina. 1986. *Introduction to classical and modern test theory*. ERIC.

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

[31] Vedant Das Swain, Manikanta D. Reddy, Kari Anne Nies, Louis Tay, Munmun De Choudhury, and Gregory D. Abowd. 2019. Birds of a Feather Clock Together: A Study of Person–Organization Fit Through Latent Activity Routines. *Proc. ACM Hum.-Comput. Interact* CSCW (2019).

[32] Robyn M Dawes and Bernard Corrigan. 1974. Linear models in decision making. *Psychological bulletin* (1974).

[33] Munmun De Choudhury and Scott Counts. 2013. Understanding affect in the workplace via social media. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 303–316.

[34] Giada Di Stefano, Francesca Gino, Gary P Pisano, and Bradley R Staats. 2016. Making experience count: The role of reflection in individual learning. *Harvard Business School NOM Unit Working Paper* 14-093 (2016), 14–093.

[35] Richard A Dienstbier. 1989. Arousal and physiological toughness: implications for mental and physical health. *Psychological review* 96, 1 (1989), 84.

[36] Joan DiMicco, David R Millen, Werner Geyer, Casey Dugan, Beth Brownholtz, and Michael Muller. 2008. Motivations for social networking at work. In *Proc. CSCW*.

[37] Joan Morris DiMicco, Werner Geyer, David R Millen, Casey Dugan, and Beth Brownholtz. 2009. People sensemaking and relationship building on an enterprise social network site. In *2009 42nd Hawaii International Conference on System Sciences*. IEEE, 1–10.

[38] Judith S Donath et al. 1999. Identity and deception in the virtual community. *Communities in cyberspace* (1999).

[39] Carol S Dweck. 2009. Mindsets: Developing talent through a growth mindset. *Olympic Coach* 21, 1 (2009), 4–7.

[40] Kate Ehrlich and N Sadat Shami. 2010. Microblogging inside and outside the workplace. In *Fourth International AAAI Conference on Weblogs and Social Media*.

[41] Amir Erez, Vilmos F Misangyi, Diane E Johnson, Marcie A LePine, and Kent C Halverson. 2008. Stirring the hearts of followers: charismatic leadership as the transferal of affect. *Journal of applied psychology* 93, 3 (2008), 602.

[42] Rosta Farzan, Joan M DiMicco, David R Millen, Casey Dugan, Werner Geyer, and Elizabeth A Brownholtz. 2008. Results from deploying a participation incentive mechanism within the enterprise. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 563–572.

[43] Franco Fiordelisi and Ornella Ricci. 2014. Corporate culture and CEO turnover. *Journal of Corporate Finance* 28 (2014).

[44] James H Fowler and Nicholas A Christakis. 2008. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *Bmj* 337 (2008), a2338.

[45] Suzy Fox, Paul E Spector, Angeline Goh, Kari Bruursema, and Stacey R Kessler. 2012. The deviant citizen: Measuring potential positive relations between counterproductive work behaviour and organizational citizenship behaviour. *Journal of Occupational and Organizational Psychology* 85, 1 (2012), 199–220.

[46] Bruno S Frey and Margit Osterloh. 2001. *Successful management by motivation: Balancing intrinsic and extrinsic incentives*. Springer Science & Business Media.

[47] Yitzhak Fried, Haim Ailan Ben-David, Robert B Tiegs, Naftali Avital, and Uri Yeverechyahu. 1998. The interactive effect of role conflict and role ambiguity on job performance. *Journal of occupational and organizational psychology* (1998).

[48] Adrian Furnham, Andreas Eracleous, and Tomas Chamorro-Premuzic. 2009. Personality, motivation and job satisfaction: Hertzberg meets the Big Five. *Journal of managerial psychology* 24, 8 (2009), 765–779.

[49] Alfonso Gambardella, Claudio Panico, and Giovanni Valentini. 2015. Strategic incentives to human capital. *Strategic Management Journal* 36, 1 (2015), 37–52.

[50] Werner Geyer, Casey Dugan, Joan DiMicco, David R Millen, Beth Brownholtz, and Michael Muller. 2008. Use and reuse of shared lists as a social content type. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1545–1554.

[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] Eric Gilbert. 2012. Phrases that signal workplace hierarchy. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1037–1046.

[53] Erving Goffman. 1981. *Forms of talk*. University of Pennsylvania Press.

[54] Erving Goffman et al. 1959. The presentation of self in everyday life. (1959).

[55] Jeffrey L Goodie, Kevin T Larkin, and Scott Schauss. 2000. Validation of Polar heart rate monitor for assessing heart rate during physical and mental stress. *Journal of Psychophysiology* 14, 3 (2000), 159.

[56] Adam M Grant, Justin M Berg, and Daniel M Cable. 2014. Job titles as identity badges: How self-reflective titles can reduce emotional exhaustion. *Academy of Management journal* 57, 4 (2014), 1201–1225.

[57] Anne E Green, Yuxin Li, David Owen, and Maria De Hoyos. 2012. Inequalities in use of the Internet for job search: similarities and contrasts by economic status in Great Britain. *Environment and Planning A* 44, 10 (2012), 2344–2358.

[58] Mark A Griffin, Andrew Neal, and Sharon K Parker. 2007. A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of management journal* 50, 2 (2007), 327–347.

[59] Xing Sam Gu and Paul R Rosenbaum. 1993. Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics* 2, 4 (1993), 405–420.

[60] Jamie Guillory and Jeffrey T Hancock. 2012. The effect of Linkedin on deception in resumes. *Cyberpsychology, Behavior, and Social Networking* 15, 3 (2012), 135–140.

[61] Jamie Guillory, Jason Spiegel, Molly Drislane, Benjamin Weiss, Walter Donner, and Jeffrey Hancock. 2011. Upset now?: emotion contagion in distributed groups. In *Proc. CHI*. ACM.

[62] Anil K Gupta and Vijay Govindarajan. 2002. Cultivating a global mindset. *Academy of Management Perspectives* 16, 1 (2002), 116–126.

[63] Nina Haferkamp, Sabrina C Eimler, Anna-Margarita Papadakis, and Jana Vanessa Kruck. 2012. Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking* 15, 2 (2012), 91–98.

[64] Juliane Hellhammer and Melanie Schubert. 2012. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology* 37, 1 (2012), 119–124.

[65] Herbert G Heneman. 1974. Comparisons of self-and superior ratings of managerial performance. *Journal of Applied Psychology* 59, 5 (1974), 638.

[66] Louis Hickman, Koustuv Saha, Munmun De Choudhury, and Louis Tay. 2019. Automated tracking of components of job satisfaction via text mining of twitter data. In *ML Symposium, SIOP*.

[67] Joseph M Hilbe. 2011. *Negative binomial regression*. Cambridge University Press.

[68] 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.

[69] Robyn Huff-Eibl, Jeanne F Voyles, and Michael M Brewer. 2011. Competency-based hiring, job description, and performance goals: The value of an integrated system. *Journal of Library Administration* 51, 7-8 (2011), 673–691.

[70] John E Hunter and Hannah Rothstein Hirsh. 1987. Applications of meta-analysis. (1987).

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

[72] Susan E Jackson and Randall S Schuler. 1985. A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational behavior and human decision processes* 36, 1 (1985), 16–78.

[73] Willow S Jacobson and Shannon Howle Tufts. 2013. To post or not to post: Employee rights and social media. *Review of public personnel administration* 33, 1 (2013), 84–107.
[74] Michal Jacovi, Ido Guy, Shiri Kremer-Davidson, Sara Porat, and Netta Aizenbud-Reshef. 2014. The perception of others: inferring reputation from social media in the enterprise. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 756–766.
[75] Shagun Jhaver, Justin Cranshaw, and Scott Counts. 2019. Measuring Professional Skill Development in U.S. Cities Using Internet Search Queries. In *ICWSM*.
[76] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online harassment and content moderation: The case of blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 2 (2018), 12.
[77] M Kahn, NP Barney, RM Briggs, KJ Bloch, and MR Allansmith. 1990. Penetrating the conjunctival barrier. The role of molecular weight. *Investigative ophthalmology & visual science* 31, 2 (1990), 258–261.
[78] Robert L Kahn and Philippe Byosiere. 1992. Stress in organizations. (1992).
[79] Robert L Kahn, Donald M Wolfe, Robert P Quinn, J Diedrick Snoek, and Robert A Rosenthal. 1964. Organizational stress: Studies in role conflict and ambiguity. (1964).
[80] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. 2012. Talking in circles: selective sharing in google+. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1065–1074.
[81] Lynda A King and Daniel W King. 1990. Role conflict and role ambiguity: A critical assessment of construct validity. *Psychological Bulletin* 107, 1 (1990), 48.
[82] Nadav Klein, Ioana Marinescu, Andrew Chamberlain, and Morgan Smart. 2018. Online Reviews Are Biased. HereâĂŹs How to Fix Them. *Harvard Business Review (https://goo. gl/CZSa5B)* (2018).
[83] Donald E Klingner. 1979. When the traditional job description is not enough. *Personnel journal* (1979).
[84] Tanja Koch, Charlene Gerber, and Jeremias J de Klerk. 2018. The impact of social media on recruitment: Are you LinkedIn? *SA Journal of Human Resource Management* 16, 1 (2018), 1–14.
[85] Keely Kolmes. 2012. Social media in the future of professional psychology. *Professional Psychology: Research and Practice* 43, 6 (2012), 606.
[86] Jaana Kuoppala, Anne Lamminpää, Juha Liira, and Harri Vainio. 2008. Leadership, job well-being, and health effectsâĂŤa systematic review and a meta-analysis. *Journal of occupational and environmental medicine* 50, 8 (2008).
[87] Nicholas Ladany and Myrna L Friedlander. 1995. The relationship between the supervisory working alliance and trainees' experience of role conflict and role ambiguity. *Counselor Education and supervision* 34, 3 (1995), 220–231.
[88] Richard S Lazarus. 1995. Psychological stress in the workplace. *Occupational stress: A handbook* 1 (1995), 3–14.
[89] Jongseo Lee and Juyoung Kang. 2017. A Study on Job Satisfaction Factors in Retention and Turnover Groups using Dominance Analysis and LDA Topic Modeling with Employee Reviews on Glassdoor. com. (2017).
[90] Lu Luo. 1999. Work motivation, job stress and employees' well-being. *Journal of applied management studies* 8 (1999).
[91] 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.
[92] William H Macey and Benjamin Schneider. 2008. The meaning of employee engagement. *Industrial and organizational Psychology* 1, 1 (2008), 3–30.
[93] James Manyika, Susan Lund, Michael Chui, Jacques Bughin, Jonathan Woetzel, Parul Batra, Ryan Ko, and Saurabh Sanghvi. 2017. Jobs lost, jobs gained: Workforce transitions in a time of automation. *McKinsey Global Institute* (2017).
[94] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proc. CHI*. ACM, 3025–3034.
[95] Alice E Marwick. 2017. Entrepreneurial Subjects: Venturing from Alley to Valley. *International Journal of Communication (19328036)* 11 (2017).
[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] Rebecca C Masson, Mark A Royal, Tom G Agnew, and Saul Fine. 2008. Leveraging employee engagement: The practical implications. *Industrial and Organizational Psychology* 1, 1 (2008), 56–59.
[98] 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. D'Mello, Anind K. Dey, Ge Gao, Krithika Jagannath, Kaifeng Jiang, Suwen Lin, Qiang Liu, Gloria Mark, Gonzalo J. Martinez, Kizito Masaba, Shayan Mirjafari, Edward Moskal, Raghu Mulukutla, Kari Nies, Manikanta D. Reddy, Pablo Robles-Granda, Koustuv Saha, Anusha Sirigiri, and Aaron Striegel. 2019. The Tesserae Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. https://doi.org/10.1145/3290607.3299041
[99] Joseph E McGrath. 1976. Stress and behavior in organization. Handbook of industrial and organizational psychology. *Chicago: RandMcnall* (1976).
[100] David E McNabb and F Thomas Sepic. 1995. Culture, climate, and total quality management: Measuring readiness for change. *Public Productivity & Management Review* (1995), 369–385.
[101] Robert K Merton. 1968. The Matthew effect in science: The reward and communication systems of science are considered. *Science* 159, 3810 (1968), 56–63.

[102] Hugh Miller. 1995. The presentation of self in electronic life: Goffman on the Internet. In *Embodied knowledge and virtual space conference*, Vol. 9.
[103] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T Campbell, Nitesh V Chawla, Munmun De Choudhury, et al. 2019. Differentiating Higher and Lower Job Performers in the Workplace using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2019), 1–23.
[104] Tanushree Mitra and Eric Gilbert. 2012. Have you heard?: How gossip flows through workplace email. In *ICWSM*.
[105] Tanushree Mitra, Michael Muller, N Sadat Shami, Abbas Golestani, and Mikhil Masli. 2017. Spread of Employee Engagement in a Large Organizational Network: A Longitudinal Analysis. *PACM HCI* CSCW (2017).
[106] Stephan J Motowidlo and Harrison J Kell. 2012. Job performance. *Handbook of Psychology, Second Edition* 12 (2012).
[107] David A Nadler, G Douglas Jenkins, Cortlandt Cammann, and Edward E Lawler. 1975. The Michigan organizational assessment package: Progress report II. *Ann Arbor: Institute for Social Research, University of Michigan* (1975).
[108] Raymond A Noe, John R Hollenbeck, Barry Gerhart, and Patrick M Wright. 2017. *Human resource management: Gaining a competitive advantage*. McGraw-Hill Education New York, NY.
[109] Robert M Ob́rien. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity* 41, 5 (2007), 673–690.
[110] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.
[111] Dennis W Organ. 1988. *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books/DC Heath and Com.
[112] Jone L Pearce. 1981. Bringing some clarity to role ambiguity research. *Academy of Management Review* (1981).
[113] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
[114] Pew. 2018. pewinternet.org/fact-sheet/social-media. Accessed: 2018-04-18.
[115] Bruce Louis Rich, Jeffrey A Lepine, and Eean R Crawford. 2010. Job engagement: Antecedents and effects on job performance. *Academy of management journal* 53, 3 (2010), 617–635.
[116] Kelsey-Jo Ritter, Russell A Matthews, Michael T Ford, and Alexandra A Henderson. 2016. Understanding role stressors and job satisfaction over time using adaptation theory. *Journal of Applied Psychology* 101, 12 (2016), 1655.
[117] John R Rizzo, Robert J House, and Sidney I Lirtzman. 1970. Role conflict and ambiguity in complex organizations. *Administrative science quarterly* (1970), 150–163.
[118] Maria Rotundo and Paul R Sackett. 2002. The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of applied psychology* (2002).
[119] Koustuv Saha, Ayse Elvan Bayraktaraglu, Andrew Campbell, Nitesh V Chawla, Munmun De Choudhury, et al. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In *CHI Ext. Abstracts*.
[120] Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 92 (Dec. 2017), 27 pages.
[121] Koustuv Saha, Manikanta D Reddy, Vedant Das Swain, Julie M Gregg, Ted Grover, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, et al. 2019. Imputing Missing Social Media Data Stream in Multisensor Studies of Human Behavior. In *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII 2019)*.
[122] Koustuv Saha, Manikanta D Reddy, and Munmun De Choudhury. 2019. JobLex: A Lexico-Semantic Knowledgebase of Occupational Information Descriptors. In *SocInfo*.
[123] 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 *Proc. ICWSM*.
[124] 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*.
[125] Alan M Saks. 2006. Antecedents and consequences of employee engagement. *Journal of managerial psychology* (2006).
[126] Marcela Sapone. 2019. medium.com/@MsSapone/why-titles-will-kill-your-startup-58402c5b6954. Accessed: 2019-03-26.
[127] Susanne Schmidt, Ulrike Roesler, Talin Kusserow, and Renate Rau. 2014. Uncertainty in the workplace: examining role ambiguity and role conflict, and their link to depression – a meta-analysis. *European Journal of Work and Organizational Psychology* 23, 1 (2014), 91–106.
[128] Neal Schmitt, Jose M Cortina, Michael J Ingerick, and Darin Wiechmann. 2003. Personnel selection and employee performance. *Handbook of psychology: Industrial and organizational psychology* 12 (2003), 77–105.
[129] W Joel Schneider and Kevin S McGrew. 2012. The Cattell-Horn-Carroll model of intelligence. (2012).
[130] N Sadat Shami, Michael Muller, Aditya Pal, Mikhil Masli, and Werner Geyer. 2015. Inferring employee engagement from social media. In *Proc. CHI*.
[131] N Sadat Shami, Jeffrey Nichols, and Jilin Chen. 2014. Social media participation and performance at work: a longitudinal study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 115–118.
[132] N Sadat Shami, Jiang Yang, Laura Panc, Casey Dugan, Tristan Ratchford, Jamie C Rasmussen, Yannick M Assogba, Tal Steier, Todd Soule, Stela Lupushor, et al. 2014. Understanding employee social media chatter with enterprise social pulse. In *Proc. CSCW*.

[133] Walter C Shipley. 2009. *Shipley-2: manual.* WPS.
[134] Jagdip Singh. 1993. Facets , Ambiguity :. *The Journal of Marketing* 57, 2 (1993), 11–31.
[135] Meredith M Skeels and Jonathan Grudin. 2009. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. In *Proc. GROUP*. ACM.
[136] Carlla S Smith, John Tisak, and Robert A Schmieder. 1993. The measurement properties of the role conflict and role ambiguity scales: A review and extension of the empirical research. *Journal of organizational Behavior* (1993).
[137] Daniel J Solove. 2007. I've got nothing to hide and other misunderstandings of privacy. *San Diego L. Rev.* (2007).
[138] 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.
[139] Sherny E Sullivan and Rabi S Bhagat. 1992. Organizational stress, job satisfaction and job performance: where do we go from here? *Journal of management* 18, 2 (1992), 353–374.
[140] Roshan Sumbaly, Jay Kreps, and Sam Shah. 2013. The big data ecosystem at linkedin. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 1125–1134.
[141] Prasanna Tambe and Lorin M Hitt. 2012. Now IT's personal: Offshoring and the shifting skill composition of the US information technology workforce. *Management Science* 58, 4 (2012), 678–695.
[142] 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.
[143] James D Thompson. 2017. *Organizations in action: Social science bases of administrative theory*. Routledge.
[144] Sigal Tifferet and Iris Vilnai-Yavetz. 2018. Self-presentation in LinkedIn portraits: common features, gender, and occupational differences. *Computers in Human Behavior* 80 (2018), 33–48.
[145] Allison S Troy, Frank H Wilhelm, Amanda J Shallcross, and Iris B Mauss. 2010. Seeing the silver lining: cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion* 10, 6 (2010), 783.
[146] Zeynep Tufekci. 2008. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society* 28, 1 (2008), 20–36.
[147] Josh Tyler. 2015. The Myth of the Ninja Rockstar Developer. In *Building Great Software Engineering Teams*.
[148] Sonja Utz. 2016. Is LinkedIn making you more successful? The informational benefits derived from public social media. *New Media & Society* 18, 11 (2016), 2685–2702.
[149] Sonja Utz and Johannes Breuer. 2019. The Relationship Between Networking, LinkedIn Use, and Retrieving Informational Benefits. *Cyberpsychology, Behavior, and Social Networking* (2019).
[150] Niels van de Ven, Aniek Bogaert, Alec Serlie, Mark J Brandt, and Jaap JA Denissen. 2017. Personality perception based on LinkedIn profiles. *Journal of Managerial Psychology* 32, 6 (2017), 418–429.
[151] José Van Dijck. 2013. 'You have one identity': Performing the self on Facebook and LinkedIn. *Media, culture & society* 35, 2 (2013), 199–215.
[152] Olivier Van Reeth, Laurence Weibel, Karine Spiegel, Rachel Leproult, C Dugovic, and Stefania Maccari. 2000. Physiology of sleep (review)–interactions between stress and sleep: from basic research to clinical situations. *Sleep Medicine Reviews* 4, 2 (2000), 201–219.
[153] Mary Van Sell, Arthur P. Brief, and Randall S. Schuler. 1981. Role Conflict and Role Ambiguity: Integration of the Literature and Directions for Future Research. *Human Relations* 34, 1 (1981), 43–71.
[154] Chockalingam Viswesvaran and Deniz S Ones. 2000. Perspectives on models of job performance. *International Journal of Selection and Assessment* 8, 4 (2000), 216–226.
[155] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 3–14.
[156] Jack W Wiley, Brenda J Kowske, and Anne E Herman. 2010. Developing and validating a global model of employee engagement. *Handbook of Employee Engagement: Perspectives, Issues, Research and Practice* (2010).
[157] Larry J Williams and Stella E Anderson. 1991. Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of management* 17, 3 (1991), 601–617.
[158] Thomas A Wright and Douglas G Bonett. 2007. Job satisfaction and psychological well-being as nonadditive predictors of workplace turnover. *Journal of management* 33, 2 (2007), 141–160.
[159] Thomas A Wright and Russell Cropanzano. 2000. Psychological well-being and job satisfaction as predictors of job performance. *Journal of occupational health psychology* 5, 1 (2000), 84.
[160] Seokhwa Yun, Riki Takeuchi, and Wei Liu. 2007. Employee self-enhancement motives and job performance behaviors: Investigating the moderating effects of employee role ambiguity and managerial perceptions of employee commitment. *Journal of Applied Psychology* 92, 3 (2007), 745.
[161] Todd R Zenger. 1994. Explaining organizational diseconomies of scale in R&D: Agency problems and the allocation of engineering talent, ideas, and effort by firm size. *Management science* 40, 6 (1994), 708–729.
[162] Todd R Zenger and Sergio G Lazzarini. 2004. Compensating for innovation: Do small firms offer high-powered incentives that lure talent and motivate effort? *Managerial and Decision Economics* 25, 6-7 (2004), 329–345.

[163] Hui Zhang, Munmun De Choudhury, and Jonathan Grudin. 2014. Creepy but inevitable?: the evolution of social networking. In *Proc. CSCW*. ACM.
[164] Julie Zide, Ben Elman, and Comila Shahani-Denning. 2014. LinkedIn and recruitment: How profiles differ across occupations. *Employee Relations* 36, 5 (2014), 583–604.

Received April 2019; revised June 2019; accepted August 2019

Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities

Koustuv Saha
Georgia Tech
koustuv.saha@gatech.edu

Amit Sharma
Microsoft Research India
amshar@microsoft.com

Abstract

Online mental health communities enable people to seek and provide support, and growing evidence shows the efficacy of community participation to cope with mental health distress. However, what factors of peer support lead to favorable psychosocial outcomes for individuals is less clear. Using a dataset of over 300K posts by ∼39K individuals on an online community TalkLife, we present a study to investigate the effect of several factors, such as adaptability, diversity, immediacy, and the nature of support. Unlike typical causal studies that focus on the effect of each treatment, we focus on the outcome and address the *reverse* causal question of identifying treatments that may have led to the outcome, drawing on case-control studies in epidemiology. Specifically, we define the outcome as an aggregate of affective, behavioral, and cognitive psychosocial change and identify *Case* (most improved) and *Control* (least improved) cohorts of individuals. Considering responses from peers as treatments, we evaluate the differences in the responses received by *Case* and *Control*, per matched clusters of similar individuals. We find that effective support includes complex language factors such as diversity, adaptability, and style, but simple indicators such as quantity and immediacy are not causally relevant. Our work bears methodological and design implications for online mental health platforms, and has the potential to guide suggestive interventions for peer supporters on these platforms.

1 Introduction

Online Mental Health Communities (OMHCs) are dedicated online support platforms aimed at aiding individuals to share, discuss and solicit information and support related to mental health. In many ways, OMHCs function like the online analog of support groups (Potts 2005). Anonymity and social connectedness in OMHCs help individuals overcome stigma and make candid self-disclosures about their mental health concerns (Andalibi et al. 2016). Examples of OMHCs include mental health subreddits on Reddit, condition-specific discussion forums on 7Cups, and social network-based interactions on Talklife (Pruksachatkun et al. 2019). OMHCs help individuals draw psychosocial benefits that help them cope with their mental health struggles (Love et al. 2012).

Given the growing popularity of OMHCs, research has studied various aspects of participation in these communities and how they may lead to better psychosocial outcomes for individuals. Extending evidence from experiments that demonstrate the efficacy of online support (Winzelberg et al. 2003), studies on Reddit and TalkLife find that they offer a thriving, global community for people to talk about their mental health (Pendse et al. 2019). They provide a fine-grained data source to understand how people express mental health distress and support each other in the real world, such as shifts in suicidal ideation (De Choudhury et al. 2016).

However, relatively little attention has been directed on the peer supporters on such platforms and how they can be more effective at providing support. A natural question to ask is *what kinds of supportive behavior leads to better outcomes for individuals receiving support*. Identifying support characteristics in responses and discussions that lead to positive psychosocial outcomes can yield insights on the best strategies of providing support, complementing work in psychotherapy literature (Norcross and Lambert 2018). Further, by focusing on natural conversations *in situ*, these insights can help OMHC owners design recommendations for their members to make more effective supportive responses.

To investigate the factors that contribute to effective support, we adopt the "case-control" study design from epidemiology (Schulz and Grimes 2002). The idea is to identify individuals who have had positive outcomes (*Case* group) and then retrospectively compare their characteristics with a similar *Control* group of individuals. Specifically, we identify people who have had long-term positive psychosocial changes and compare the characteristics of responses they received to that received by those who did not have such positive changes. In the language of causal inference, each response from a peer supporter can be considered as an intervention for an individual, and we are interested to find the characteristics of interventions that lead to the maximum positive change. Using the symmetry of the back-door method (Pearl 2009), we argue that looking for interventions that vary significantly between case and control groups translate to finding interventions with causal effects on the outcome.

Specifically, we work with a longitudinal dataset of ∼39K individuals on TalkLife, an online mental health platform. We quantify their psychosocial outcomes as an aggregate measure of affective, behavioral, and cognitive outcomes. On the basis of psychosocial change from when they joined to the present, we obtain two separate cohorts of individuals — psychosocially most (*Case*) and least (*Control*) improved. We compare the differences in responses across a range of characteristics drawn on psychotherapy literature like adaptability, immediacy, diversity, emotionality, language style,

and nature of support. Confirming past work, we find complex linguistic attributes such as adaptability, diversity, and style are significant factors for driving positive psychosocial change. In particular, factors related to adaptability such as topical congruence and linguistic accommodation have the highest difference between *Case* and *Control* groups. Somewhat surprisingly, the average length of responses has a substantial positive effect towards driving people to better outcomes, possibly as a proxy for the linguistic factors described above. Other simple factors, however, such as number of responses and immediacy of receiving a response do not have significant differences between *Case* and *Control*.

Compared to previous work by De Choudhury and Kıcıman (2017) estimating effects of using specific phrases in responses, our work has an advantage whenever one is interested in analyzing continuous treatments and *finding* the interventions that lead to desired outcomes. This is because most *forward* causal inference methods (Gelman and Imbens 2013) require binarization of treatment variables. In contrast, case-control methods avoid apriori binarization of complex treatments and estimate the differences in treatment instead. Especially in OMHCs where everyone typically receive responses, our proposed method is useful to determine the necessary *dosage* increase of support treatments that can increase the likelihood of positive outcomes (Hernan and Robins 2010). There is also a computational advantage. In forward causal inference methods (Rubin 2005), one may estimate a separate propensity score model for each treatment whereas our case-control method allows estimating the differences in multiple treatments at once. We discuss the methodological and practical implications of our work towards improving support in OMHCs through recommendation-based interventions for peer supporters.

Privacy, Ethics, and Disclosure. This paper uses sourced data (licensed and consented) from TalkLife. Our work is in collaboration with TalkLife, and given the sensitivity of our work, we are committed to securing the privacy of the individuals. The dataset was accessed through secured databases with necessary privacy and ethical protocols in place, and the dataset was de-identified and no personally identifiable information was used. This paper only reports aggregated measures to prevent traceability and identifiability of individuals on the platform. Even accounting for the benefits, we recognize the potential misuses, risks, and ethical consequences involved with this kind of research, which we elaborate in Discussion. This work is approved by the Institutional Review Board at Microsoft Research.

2 Background and Related Work

Effective Psychotherapeutic Interventions What constitutes effective counseling and psychotherapeutic strategies has interested researchers and practitioners for a long-term now (Labov and Fanshel 1977), and treatments and therapeutic methods constantly evolve and advance over time. Lambert and Barley (2001) formulated four areas that influence a care-seekers' outcome in psychotherapeutic settings – extratherapeutic factors, expectancy effects, specific therapy techniques, and common factors. Among these, common factors include empathy, warmth, congruence, therapeutic alliance, are found to be most highly correlated with the outcomes. In another work, Norcross and Lambert (2018) conducted a meta-analysis on the effectiveness of several elements of psychotherapeutic relationships.

In the area of online technology aided and mediated mental health interventions, Cavanagh et al. (2018) showed the efficacy of computer-mediated psychotherapy towards positive clinical outcome. Although its efficacy is yet to be established and results are mixed (Rollman et al. 2018), researchers have stressed the importance of social media as a mental health intervention platform (Chikersal et al. 2020, Ernala et al. 2017, Merolli et al. 2013, Yoo and De Choudhury 2019). Relatedly, Dinakar et al. (2015) used computational linguistics and machine learning to improve crisis counseling and interventions, Haberstroh et al. (2007) studied online counseling experiences, and Althoff et al. (2016) studied effectiveness of counseling language including adaptability, creativity, and perspective change. Our work draws upon these prior works to evaluate what factors help in desirable psychosocial outcomes in "online psychotherapeutic setting", where individuals seek and share mental health related support.

Support in Online Mental Health Communities (OMHCs) With the widespread use of social media-based technologies, OMHCs are becoming increasingly popular. Originally ideated as online analogs of support groups, individuals in these communities share and seek support related to sensitive mental health concerns faced by themselves or their loved ones (De Choudhury and De 2014; Huh 2015; Saha et al. 2019a; 2020; Sharma and De Choudhury 2018; Kummervold et al. 2002). Prior work studied how anonymity, engagement, social capital, and social connectedness help in candid self-disclosure and seeking mental health support (Andalibi et al. 2016; Ernala et al. 2018). In addition, online social support is known to build interpersonal relationships, and to improve psychological wellbeing, self-esteem, satisfaction, and reciprocity (Steinfield et al. 2008; Oh et al. 2013).

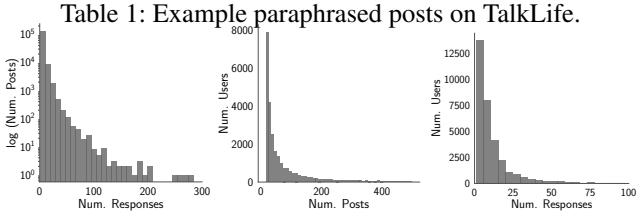
For specialized OMHC platforms such as Talklife or 7Cups, recent studies have examined positive outcomes over a sample of users (Baumel et al. 2018) or over a single thread or bursts of conversation (Kushner and Sharma 2020, Pruksachatkun et al. 2019; Pendse et al. 2019). We extend this research by focusing on long-term changes in one's mental health over a large sample of individuals and retrospectively finding the most relevant causes.

Causal Inference Studies on Observational Data The gold-standard approach to establish causality is via a randomized controlled trial. In early work, such trials were conducted to assess the efficacy of online support communities of breast cancer (Winzelberg et al. 2003). However, trials are not always feasible due to practical and ethical concerns (Hannan 2008). As an alternative, researchers resort to observational studies. While these cannot guarantee causality, observational studies allow to investigate long-term and longitudinal data, and are especially useful to find candidate treatments for a future randomized trial when no preferred treatment is known apriori (Rubin 2005). There are two popular means of conducting observational studies — 1) cohort based, where the treatment is known and the goal is to find its causal effects on the outcome, and 2) case-control based, where the outcomes are known, and the goal is to (retrospectively) find the treatments that potentially caused the outcome.

In our related space of social media and mental health,

Erlyer i slit my hand open and forgot about it oh well
I don't get why my parents just love hurting me.
i feel very empty & exposed. [..] i hate being sick but its just in my head[..] want to stop breathing

Table 1: Example paraphrased posts on TalkLife.



(a) Resp. per post (b) Posts per user (c) Resp. per user

Figure 1: Distribution of posts and responses in study dataset.

observational studies have examined the effects of suicidal ideation (De Choudhury et al. 2016), social support (De Choudhury and Kıcıman 2017), psychiatric medications (Saha et al. 2019b), exercise (Dos Reis and Culotta 2015), alcohol use (Kıcıman et al. 2018), crisis (Saha and De Choudhury 2017), and counseling interventions (Saha et al. 2018). These studies adopted the cohort based, or prospective analysis of conditioning on treatments, and matching on similar individuals to examine the differences in the outcomes, which finally quantifies the causal effects (Olteanu et al. 2017). However, cohort-based studies may not be suitable when the goal is to rank multiple continuous treatments on their causal effect, especially when almost everyone receives variable treatment dosage, and there are no obvious way to binarize the different treatments. We therefore adopt the case-control study design, and provide a method to analyze the effects of several continuous treatments.

3 Data

The dataset of our study comes from TalkLife (Pendse et al. 2019), an online mental health discussion forum, self-describing itself as a "safe social network to get help and give help". From the standpoint of social computing interface, TalkLife functions like many other online communities and discussion forums. The community members participate via discussion threads, where each discussion thread consists of an original post (or "post" hereon), and a number of responses by community members which are typically relevant to the post in the discussion thread. We obtain data from the TalkLife platform (in collaboration with TalkLife) from August 2011 to January 2019. This data includes discussion threads, each with a single post and a number of responses by community members. We note that TalkLife responses can also follow a hierarchical nature, however our work accounts for all kinds of responses similarly, under an umbrella term of "responses". Our dataset consists of a total of over 6.5M posts (equal number of discussion threads) and 20M responses posted by over 300K users. On average, each discussion thread receives 4.44 responses (stdev.=30.60). Table 1 shows examples of four paraphrased posts on TalkLife.

Social computing platforms tend to change significantly over the years, which could also lead to changes in usage behavior and objectives of individuals. For the purposes of our study, to minimize the effects of long-term platform-level changes and aggregated use behavior, we limit our analysis to 38,977 individuals who started participating in TalkLife after January 01, 2016, continued participation (posted more than once) beyond January 01, 2018, and overall had posted at least 15 times on the platform. Given that our data lasts till January 2019, each user has been on the platform for at most three years (mean=222 days, median=142 days). Given that the TalkLife platform has stabilized over recent years, we expect a lower impact of platform-level changes during this three-year time period. This study concerns a dataset comprising of 3,184,612 posts and 23,528,159 responses with 4.72 response per post (stdev. = 54.81). Figure 1 shows the distribution in our dataset as responses per post (Figure 1(a)), posts per user (Figure 1(b)), and responses per user (Figure 1(c)).

4 Study Design and Methods

Retrospective Case-Control Design

We are interested in the effect of different kinds of supportive interventions (or treatments) on psychosocial outcomes for users on OMHCs. If interventions are well-defined (e.g., in medicine, whether a drug was prescribed), then the most common approach is to estimate the effect of each intervention separately. If T is the treatment, Y the outcome, and W represents common causes or confounders of T and Y, then the causal effect (Pearl 2009) of T on Y is represented as:

$$E[Y|do(T=1)] - E[Y|do(T=0)] = \sum_{w} E[Y|T=1,W] - \sum_{w} E[Y|T=0,W] \quad (1)$$

Effectively, the method compares outcomes for people with or without the intervention while conditioning on all confounders. However, when there are multiple candidate treatments for the same outcome of interest, it may be more appropriate to ask the *reverse* causal question. That is, rather than the effect of a treatment, we ask about the potential causes of an observed outcome. In our problem, for example, the outcome is pre-specified—psychosocial health of individuals—but treatments are not. At one level, we could consider participation on Talklife as a means to be treated. At another level, we would like to consider the effect of several characteristics of language and behavior used by peer supporters and isolate their effects. Thus, our goal is to *find* treatments that lead to a significant change in the outcome.

When the outcome is well-specified, and it distinguishes two groups of *Case* and *Control* users, we can use the following estimator for finding the interventions that causally affect the outcome, based on the case-control study design in epidemiology (Schulz and Grimes 2002):

$$\sum_{w} E[T|Y=1,W] - \sum_{w} E[T|Y=0,W] \quad (2)$$

In our work, W corresponds to covariates consisting of individual attributes. Intuitively, this equation refers to conditioning on W, and then calculating the differences between treatment for *Case* and *Control*. The estimator compares individuals with a positive outcome ($Y = 1$) with others (Y=0) and measures the difference in intervention values between the two groups, while conditioning on all known confounders. This is the so-called 'reverse" causal inference problem (Gelman and Imbens 2013). Whenever there is a significant change in T from $Y = 0$ to $Y = 1$ keeping all confounders W constant, it implies that there is a causal effect

of T on Y. Given the same outcome, we can do this analysis repeatedly for finding the treatments with the highest effects. As we will see later, the case-control analysis provides some computational benefits especially when working with continuous intervention variables and selecting a suitable cutoff to binarize them for a future randomized experiment. Further, in contrast to cohort-based causal inference studies that condition on individual treatments, our approach allows accounting for a combination of multiple treatments together on the same individuals (emulating closer to real-world settings).

The rest of this section contextualizes the above estimator in an OMHC. We operationalize the outcome on psychosocial health, determine the case and control groups, and then finally list potential interventions that we test.

Measuring the Outcome: Psychosocial Health

Towards our research objective of understanding effective psychotherapeutic interventions, we first operationalize "improvement in mental health outcome" as observed on TalkLife. Researchers have argued on what constitutes improvement and success in psychotherapeutic, psychological, and psychiatric care (Perkins 2001), where traditionally symptom reduction has been considered to be the improvement in quality of life. Generally speaking, "psychosocial health" is considered as an appropriate terminology that not only encompasses both psychological and social wellbeing, but also places the locus of health in the individual by including social wellbeing in the form of social adjustment and environmental response (Larson 1996). We situate our work on the impacts of social media based interventions on one's psychosocial health and wellbeing (Merolli et al. 2013). Given that psychosocial health is a complex construct and there is no easy means to quantify it, we adopt a conservative definition of psychosocial health based on observed behavior on the platform. As a user's posting behavior is our only available data, we draw upon prior work that operationalize therapeutic responses in online mental health communities and social media, grounded on psychology, psychiatry, and expressive writing literature (Ernala et al. 2017, Saha et al. 2018). We broadly group these observed psychosocial outcomes in three categories — *affective*, *behavioral*, and *cognitive* outcomes (Breckler 1984), and then aggregate them to construct a single outcome metric.

Affective Outcomes Simplistically, affect refers to an emotional response, and affective behavior is indicative of one's psychological wellbeing. Because social media posts are written in a self-motivated and self-initiated fashion, language is a strong means to infer affective psychosocial health. To measure this, we use the following:

Affective Words. We use the psycholinguistic lexicon, Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2003) to obtain proportion of affective (positive and negative affect) keywords per user. This draws upon expressive writing literature which associate language with therapeutic symptoms. Increased use of positive affect and decreased use of negative affect words correspond with psychosocial improvement.

Language Indicative of Mental Health Symptomatic Outcomes. To identify the presence of mental health concerns, prior work built machine learning classifiers of social media language indicative of depression, anxiety, stress, suicidal ideation, and psychosis (Saha et al. 2019b). These are n-gram (n=1,2,3) based binary SVM models. For training these classifiers, the positive class comes from domain dependent data on Reddit (*r/depression*, *r/anxiety*, *r/stress*, *r/SuicideWatch*, and *r/psychosis* subreddits for the corresponding classifiers), and the negative training examples come from a set of random non-mental health related content from Reddit. Similar to Saha et al. (2017; 2019b), we conduct linguistic equivalence test — cosine similarity of the word embedding representations of top 500 n-grams in the Reddit and Talklife datasets shows a high similarity of 0.92. This entails very similar transfer datasets, and similar performance, given that the classifiers have performed reasonably well when transferred on other social media datasets (Saha et al. 2019b). Using these classifiers, we obtain the aggregated proportion of posts that express mental health concerns corresponding to each TalkLife user. That is, lower the proportion of posts expressing mental health concerns, better is one's psychosocial health.

Behavioral Outcomes Behavioral psychosocial health consists of an individual's overt actions, behavioral intentions, and verbal statements regarding behavior (Breckler 1984). Behaviors such as changes in social functioning and shift of interests could be indicative of an individual's changing psychological trajectory (Saha et al. 2018, Guntuku et al. 2019). To quantify behavioral psychosocial outcomes, we obtain three attributes on an individual's behavior on the platform. The first of these is *activity*, or the frequency of participation on the platform – this is quantified as the number of posts per day for every individual. The second is *interactivity*, or how interactive an individual is — this is quantified as the ratio of the number of responses (to others' posts) to the number of self-posts. This essentially quantifies an individual's behavior of providing support compared to seeking support. The final one is *interaction diversity*, or the topical diversity of discussions an individual engages in — each discussion thread is labeled with a particular topic (eg., relationships, family, self-harm, friends, hopes, etc.) by the original poster. These measures are directly associated with psychosocial health — an increase in these measures corresponds to an improvement in psychosocial health (Saha et al. 2018).

Cognitive Outcomes Beliefs, knowledge structures, perceptual responses, and thoughts constitute cognitive component of psychological wellbeing (Breckler 1984). Cognitive attributes is another indicator of an individual's psychological health (Bandura 1993). Drawing on psycholinguistics literature that demonstrates how the style and structure in language define one's cognitive behavior, we adopt the following measures to define cognitive psychosocial health.

Readability measures the ease with which a reader can understand a given text. We adopt the Coleman-Liau Index (CLI) which provides readability assessment based on character and word structure within a sentence, calculated as, $CLI = (0.0588L - 0.296S - 15.8)$, where L is the average number of letters per 100 words, and S is the average number of sentences per 100 words. A greater CLI measure indicates a better writing quality, and an increase of CLI indicates psychosocial improvement (Ernala et al. 2017).

Complexity and Repeatability capture one's cognitive state in the form of planning, execution, and memory (Ernala et al. 2017). We quantify complexity as the average length of words per sentence, and repeatability as the normalized

occurrence of non-unique words. While linguistic complexity has a positive association with one's psychosocial health, repeatability shares a negative association with the same.

Psycholinguistic Keywords. We use LIWC lexicon to obtain the proportion of keywords corresponding to cognition, perception, and linguistic style categories, where linguistic style keywords correspond to non-content keywords in language such as, temporal references (past, present, and future tense), lexical density and awareness (auxilliary verbs, preposition, adverbs, verbs, articles, conjunctions, inclusive, and exclusive), and interpersonal focus (1st person singular and plural, 2nd person, and 3rd person pronouns). Literature posits the importance of keywords in understanding cognitive behavior. For instance, the variations in pronoun use reflects the transformation in the way individuals think about themselves with respect to others, and the use of articles and adverbs could indicate how individuals process complex narratives (Pennebaker et al. 2003). A greater use of these keywords is associated with one's improved psychosocial state.

Overall Psychosocial Outcome After normalizing each of the above outcomes on a min-max scale of 0 to 1, we operationalize psychosocial health of an individual as a composite measure of unit-weighted and sign-adjusted average across each of the outcomes so that higher values indicate a better psychosocial health (see below). As noted before, while this cannot be argued to be perfect, we believe that by accounting for several symptomatic observable changes on the platform, such a composite measure should be theoretically correlated with the actual psychosocial health of an individual.

$$outcome = \mu(pa - na - mh\_language + activity + int.\_diversity + interactivity + readability - complexity - repeatab. + cog.\_words)$$

Determining *Case* and *Control* Individuals

To understand the effects of support, social interactions, and responses (treatment), we identify and distinguish those individuals who improved the most, and those who did not after a period of time on the platform. Adopting terminologies from epidemiological observational studies, we name these groups as *Case* (improved) and *Control* (not improved or worsened) (Schulz and Grimes 2002). Ideally the improvement should be determined based on the change in mental health state in the present from their initial state on the platform (or before they are treated). As a proxy of the initial state, we quantify an individual's baseline psychosocial health on their first n_1 posts since they joined the platform. Treatment correspond to the attributes of responses received on the next n_2 posts on the platform and the outcome as the average psychosocial health over all posts after n_2. Essentially, we draw on variable treatment effect framework (Hernan and Robins 2010), and segregate an individual's timeline of activities on the platform into pre-treatment phase, treatment phase, and post-treatment phase (see Figure 2 for a schematic overview of the segregation on an individual's timeline). We choose the number of posts by a user instead of duration on the platform since it provides a better metric for exposure to responses given high variance in people's activity.

Choice of n_1 and n_2 There is a tradeoff in choosing n_1. A smaller n_1 ensures that we capture the initial state of a user without the effects of responses, but also exposes us to high variance in estimating it. We thus vary combinations of n_1 and n_2 in different values between 2 and 8 posts, and check for consistency in our findings. We observe that our results are not sensitive to the choice of n_1 and n_2. For the ease of exposition, we first discuss and report findings with pre-treatment phase of n_1=3 and treatment phase of responses received in the next n_2=8 posts per individual. Following this, we revisit the robustness of our findings for different combinations of n_1 and n_2.

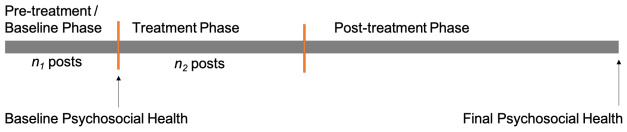
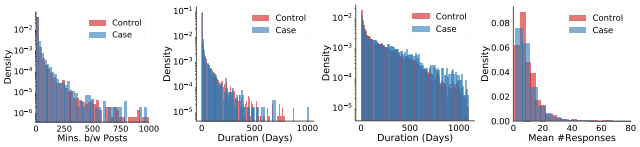


Figure 2: Schematic figure of a user's timeline in our study design. Psychosocial health is determined as an average of observed measures in the corresponding period.

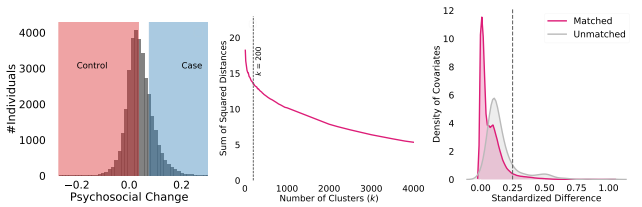


(a) (b) (c) (d)

Figure 3: Distribution in *Case* and *Control* in terms of (a) time between successive posts (b) time period in treatment phase, (c) time period in post-treatment phase, (d) number of responses (per user) received in the treatment phase.

Case and Control Individuals Within our dataset, we find that psychosocial change ranges between -0.27 and 0.36, with a mean change of 0.04 (std.=0.05). For the purposes of our study, we define *Case* to be the individuals who lie in the top 80 percentile of psychosocial change, and *Control* to be the individuals in the bottom 50 percentile of psychosocial improvement (see Figure 4a). Our choice of 80% is motivated by the idea of restricting *Case* to only people with *very good* outcomes so that we can better understand the nature of support interventions behind those changes. With this definition, we obtain 6,789 *Case* and 16,972 *Control* individuals, whom we study for our ensuing analyses.

Testing Comparability of *Case* and *Control* Before conducting causal analysis on *Case* and *Control*, we evaluate if their data is comparable. We compare the duration between successive posts made by *Case* and *Control* individuals (Figure 3a). We find that *Case* posts are separated by an average 20 minutes and *Control* posts are separated by an average 18 minutes. There is no statistical significance as per independent sample t-test and Kolmogorov-Smirnov (KS) test (p>0.1). Given that our study design rests upon a threshold on the number of posts for specifying the treatment phase (Figure 2), we test if this specification leads to any biases in length of participation time period by comparing the distribution of time period per *Case* and *Control* individual in treatment (Figure 3b) and post-treatment (Figure 3c) phases. The mean length of time in the treatment-phase is 15 days for *Case* and 16 days for *Control* individuals, and the same in the post-treatment phase is 221 days for *Case* and 204 days for *Control* individuals. For both comparisons we find



(a) Outcome Changes (b) Optimal k (c) Matching Balance

Figure 4: (a) Dist. of Psychosocial outcome of all individuals, (b) k-means clustering of individuals on covariates for several k, (c) Standardized differences following matching.

no statistical significance as per t-test and KS-test (p>0.1). Again, because we consider receiving responses as treatment, we compare if *Case* and *Control* received different number of responses overall, where we find no statistical significance as per t-test and KS-test (p>0.1). These tests provide evidence that *Case* and *Control* datasets are comparable, with minimal biases due to unaccounted measures.

Matching of Similar Individuals

We next aim to *identify the causes* of post-treatment psychosocial outcomes. We adopt a case-control framework, that conditions on the outcomes to differentiate the treatment between *Case* and *Control*. Theoretically, given two similar individuals, their likelihood to improve is similar if they were subjected to the same treatment. Thus, outcome difference is potentially caused by the differences in treatment, provided the biases due to confounders are minimized (Silber et al. 2001). We assume that all potential interventions are via responses on TalkLife, or more generally that interventions outside the platform similarly affect both *Case* and *Control*.

Covariates To reduce biases associated with confounders, the first step involves identifying a suitable set of covariates. The covariates include the exact same affective, behavioral and cognitive measures of psychosocial health, as described in the previous subsection on outcomes. However, while the outcomes are measured over posts that come after the first $n_1 + n_2$ posts, the distinction is that we compute these covariates for matching using only their first n_1 posts. In effect, we control for covariates that are baseline behavioral and psychological attributes of individuals. For each covariate, we quantify an aggregated measure per individual within their pre-treatment (first n_1) posts and responses received to them. The covariates include their pre-treatment psychosocial measures, which are: affective measures (normalized quantity of affective words and classifiers of depression, anxiety, stress, psychosis, and suicidal ideation), behavioral measures (activity, interactivity, and interaction diversity), and cognitive measures (readability, complexity, and repeatability). The covariates additionally include the top 500 n-grams (n=1,2,3) per user, and the pre-treatment average number of responses received per post per individual. The choice of covariates is motivated by prior work (Kıcıman et al. 2018, Saha et al. 2019b). We use these covariates as features in clustering similar users in our ensuing matching step.

Matching Approach To find statistically comparable individuals, we use matching. This simulates a randomized trial setting by conditioning on as many as covariates as possible (Rubin 2005). To compare against counterfactual scenarios, for those who improved (*Case*) we find their similar (matched) counterparts among the ones who did not improve (*Control*). Typically, matching methods match individuals on the basis of the likelihood of being treated, however, in our case every individual is treated (or exposed to responses), although the "dosage" of treatment measures may vary. To account for variable treatment across individuals (Hernan and Robins 2010), we match individuals in an unsupervised fashion using k-means clustering. This approach functions like a stratified matching approach (Kıcıman et al. 2018), where each cluster (or stratum) consist of matched individuals.

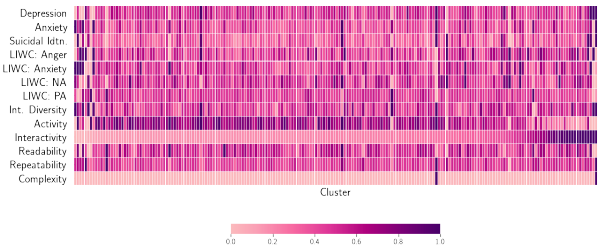


Figure 5: Heatmap showing mean values for a sample of covariates across 200 clusters. Values are rescaled using min-max scaling of 0 to 1 within a covariate.

To determine the number of clusters (k) in k-means, we use the well-adopted elbow heuristic — optimal k can be located around the greatest drop in density across clusters. Figure 4b plots the sum of squared distances of samples to the nearest cluster centroids for k varying between 1 and 5,000. Manually inspecting Figure 4b and using Kneedle algorithm (Satopaa et al. 2011), we approximate that the greatest drop (maximum curvature) occurs at around k=200, which we adopt as the number of clusters in our analysis.

After we cluster similar individuals, we drop those clusters without sufficient number of *Case* and *Control* users as these clusters could lead to biased findings (Kıcıman et al. 2018). Using a threshold of at least 10 *Case* and 10 *Control* user per cluster, we obtain 181 usable clusters that contain 6,758 *Case* and 16,920 *Control* individuals in total — together 99.6% of the Case-Control users that we identified. Each cluster essentially contains similar *Case* and *Control* individuals conditioned on pre-treatment covariates (psychosocial health and language on the platform). For better interpretation, we label the clusters by ranking on average interactivity of cluster members, i.e., those with greater interactivity are more likely to be placed at a higher value cluster. For a sample of interpretable covariates, Figure 5 shows differences between the clusters on multiple dimensions — for e.g., consider the pair of first two clusters, while the first cluster shows higher anxiety and higher interaction diversity, the second cluster shows higher suicidal ideation, anger, and activity.

Evaluating Balance The purpose of matching is to ensure that confounders are minimized to the maximum extent due to individual differences, and to help conduct like-for-like comparisons. We evaluate the balance of the covariates using standardized mean differences (SMDs) across covariates in *Case* and *Control* groups. Two groups are considered to be balanced if the covariates reveal SMD lower than 0.25 (Kıcıman et al. 2018)— a condition not fulfilled in only 0.15% cases (364 out of 207,844 covariate-cluster combinations). Further, a significant drop (57%) in mean SMD from

0.16 (sd=0.13) in the unmatched dataset to 0.07 (sd=0.08) in the matched dataset, indicates a good balance (Figure 4c).

5 Potential Causes of Psychosocial Outcomes

We now study the factors that potentially contribute to psychosocial changes. Below we list potential treatment measures that are based on the literature as contributors to psychosocial change. We hypothesize that these factors, both implicit and explicit in the responses contribute to an individual's psychosocial outcomes. Some of these factors are likely to be correlated among each other, aligning with the fact that "causes" of ones' psychosocial outcomes are inherently coarse and complex combination of these factors.

Number of Responses We hypothesize that *Case individuals received greater number of responses to their concerns*. We rationalize that receiving more responses is associated with greater social support and sense of belonging in community, which have been found to be effective in an individual's psychosocial improvement in psychology literature (Glass and Maddox 1992). To test this hypothesis, for every individual, we calculate the average number of responses received per post in the treatment phase, and compare these averages in the matched samples of *Case* and *Control* individuals.

Verbosity Along the lines of the above, we hypothesize that *Case individuals received more verbose or longer responses to their concerns*. We compare the length of responses in terms of the number of words and number of unique words received by the *Case* and *Control* individuals.

Immediacy Because immediate and sooner responses are generally recommended in the cases of mental health crisis (Flannery and Everly 2000), we hypothesize that *Case individuals received more immediate responses*. Essentially, we computed the average time to the first response received by the *Case* and *Control* individuals.

Diversity/ Creativity Drawing on the efficacy of counseling and psychotherapy styles (Althoff et al. 2016, Norcross and Lambert 2018), we hypothesize that *Case individuals received more diverse responses, in comparison to the Control individuals who received more templated and generic responses*. To examine this, we obtain the lexico-semantic diversity within the responses received by the *Case* and *Control* individuals. In particular, we use the 300-dimensional word embedding vector representations (Mikolov et al. 2013). For the responses in either of *Case* or *Control* corpus, we find their average cosine distance from the centroid of the corresponding corpus. This operationalizes the diversity in responses within *Case* and *Control* corpuses.

Emotionality We hypothesize that *Case individuals received responses that contained greater emotions and positive affirmations* (Norcross and Lambert 2018). For this, we use LIWC to obtain the normalized occurrences of affective keywords in the responses received by the *Case* and the *Control* individuals (Pennebaker, Mehl, and Niederhoffer 2003).

Adaptability We hypothesize that *Case individuals received responses that were more customized and attuned to their concerns*. This draws upon literature postulating that adaptable and linguistically accommodating responses are more effective in support than templated or generic responses (Althoff et al. 2016; De Choudhury and Kıcıman 2017). Better adaptability aids improved social feedback, solidarity, social exchanges, and reciprocated feelings of intimacy (Ferrara 1991). We examine adaptability in two measures, *topical congruence* and *linguistic style accommodation*.

Topical Congruence. Motivated by Pennebaker et al.'s (2003) work that content words are indicative of numerous psychosocial aspects, we extract the content words in responses and posts (using LIWC). We operationalize topical congruence between a response and the original post as the lexico-semantic similarity between the two, for which we obtain the cosine similarities between their word embedding representations (Das Swain et al. 2020; Tan et al. 2016).

Linguistic Style Accommodation We obtain linguistic style accommodation of each query to by using Linguistic Style Matching (Sharma and De Choudhury 2018). We compute the cosine similarity of each response and original post on the normalized occurrences of non-content or linguistic style dimensions — these are function words across the categories of articles, prepositions, pronouns, auxiliary verbs, conjunctions, adverbs, negations, etc (Pennebaker et al. 2003).

Credibility of the Responders People tend to show trust in more credible and reputable individuals in the community (Ma et al. 2019). Accordingly, we hypothesize that *Case individuals received responses from those who are experienced "care and support" givers in the community*. To measure responders' experience of providing support on the platform, we quantify their tenure (or duration of time spent), interactivity (ratio of number of responses to number of posts), and activity (number and rate of posting) on TalkLife.

Language Style of Responses Literature posits the importance of language style in effective psychotherapy (Norcross and Lambert 2018). Using personal opinion induces a sense of belonging, and also corresponds to mindful genuineness on the part of the peer-supporter. The nature of communication is a direct correlate of the complexity of language (Kolden et al. 2011). Language style can be characterized as categorical and dynamic (Pennebaker et al. 2014). Theoretically, categorical language includes approaching the world in a relatively logical, complex, and "amateur scientist" manner, and dynamic language is typically used by individuals who are more socially engaged, tell stories, and pay more attention to the world around them. We hypothesize that *"the responses received by the Case individuals is more dynamic"*. We adopt the measure of categorical-dynamic index (`CDI`) proposed by Pennebaker et al. (2014). This is a bipolar index, where higher `CDI` indicates a categorical style, and lower `CDI` indicates a dynamic or narrative style. In particular, `CDI` for a given text is quantified based on the percentage of words per style related parts of speech:

$\texttt{CDI} = (30 + \textit{article} + \textit{preposition} - \textit{personal pronoun} - \textit{impersonal pronoun} - \textit{aux. verb} - \textit{conjunction} - \textit{adverb} - \textit{negation})$

Nature of Support Past work suggests that social support is greatly effective in helping individuals cope with mental health struggles (Kummervold et al. 2002). Situated in the "Social Support Behavioral Code", two forms of support that have received theoretical and empirical attention are *emotional* and *informational* support. Emotional support corresponds to language containing empathy, encouragement, and kindness, and is considered to be most effective form of psychosocial support (Sharma and De Choudhury 2018).

	Emo. Support		Inf. Support	
Metric	**Mean**	**Max.**	**Mean**	**Max.**
Precision	0.71	0.78	0.73	0.87
Recall	0.71	0.80	0.77	0.87
F1	0.70	0.77	0.73	0.87
Accuracy	0.71	0.78	0.77	0.87
AUC	0.79	0.91	0.82	0.92

Table 2: Accuracy metrics of the Support Classifiers

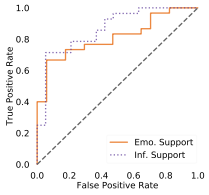


Figure 6: ROC Curve of Support Classifiers

Informational support, corresponds to providing information and advice, and is also known to be effective in positively impacting perceived empathy (Nambisan 2011). We hypothesize that *Case individuals received greater support*.

We obtain the presence of emotional (ES) and informational (IS) support in responses. We use a dataset built and expert-verified in prior work (Sharma and De Choudhury 2018) that labels supportive responses on Reddit on the degree of ES and IS. We build two binary SVM classifiers with linear kernel, either of which characterizes the degree (high/low) of ES or IS in a post. The k-fold cross validation accuracy (k=10) of ES and IS classifiers are 0.71 and 0.77 respectively (Table 2 reports accuracy metrics, Figure 6 shows ROC curves). While the classifiers are expected to perform well, better accuracy can be achieved with sophisticated models and expert annotation on TalkLife, our objective here is to leverage the feasibility of measuring nature of support in language. Note that similar to the mental health classifiers, support classifiers are also transferred from Reddit to TalkLife data, and the linguistic similarity between the two datasets ensures a reliable transfer (Section 4). We use these classifiers to machine label all responses — 15% responses contain ES and 3.3% responses contain IS, and then compare their prevalence per *Case* and *Control* group.

6 The Effect of Supportive Interventions

Per previous section, we now test the hypotheses to quantify the differences per treatment across the matched samples of *Case* and *Control* individuals. We obtain effect size (Cohen's d), and evaluate statistical significance in differences using independent sample t-test. We conduct Kolmogorov-Smirnov (KS) test which essentially tests against the null hypothesis that the distributions of treatments in the *Case* and *Control* groups are drawn from the same distribution. Table 3 summarizes these differences, which we discuss here.

Number of Responses Figure 3d shows the distribution of number of responses received by *Case* and *Control* individuals in the treatment period. While on an average, *Case* individuals receive 25% more responses than their matched *Control* individuals, we find no significant differences in the number of responses received by the matched *Case* and *Control* individuals. This suggests that individuals with similar concerns, and psychological and social attributes (because of the matching framework), are likely to receive similar number of responses. Therefore, we cannot reject the null hypothesis that the number of responses received by *Case* and *Control* individuals are from the same distribution.

Verbosity *Case* individuals receive more verbose responses than the matched *Control* individuals. This is revealed by both average length of response (t=9.08, p<0.05),

Measure	**Case**	**Control**	***d***	***t***	***KS***
Num. Responses	16.34	12.98	0.09	0.88	0.13
Verbosity (Avg. Per Response)					
Num. Words	19.98	18.07	0.96	9.08***	0.43***
Num. Unique Words	17.60	16.19	0.97	9.20***	0.48***
Immediacy (Minutes)	6.22	5.95	0.11	1.05	0.09
Diversity/Creativity	0.66	0.63	1.09	9.18***	0.46***
Emotionality (% Words)					
Anger	0.71	0.70	0.05	0.43	0.16*
Anxiety	0.33	0.31	0.15	1.41	0.12
Sadness	0.48	0.47	0.08	0.73	0.13
Neg. Affect	0.87	0.87	0.00	0.02	0.11
Pos. Affect	5.67	5.49	0.21	1.91**	0.12**
Swear	0.36	0.36	0.04	0.38	0.17*
Adaptability					
Topical Congruence	0.65	0.61	1.22	11.55***	0.50***
Linguistic Accommodation	0.80	0.61	1.38	18.14***	0.63***
Credibility of Responders					
Tenure (days)	233.93	234.49	-0.00	-0.06	0.08
Interactivity	22.46	19.39	0.32	2.93**	0.13*
Num. Posts	2987.34	2721.91	0.61	5.75***	0.40***
Posts per Day	11.73	11.01	0.15	1.42	0.10
Language Style (CDI)	3.39	4.10	-0.45	-3.74***	0.25***
Nature of Support					
Emotional	0.20	0.17	0.93	8.82***	0.36***
Informational	0.05	0.04	0.55	5.20***	0.31***

Table 3: Summary of differences in responses received by *Case* and *Control* individuals. We report average occurrences across matched clusters, effect size (Cohen's d), independent sample t-statistic, and KS-statistic. Rows with significant differences are shaded in grey, p-values are reported after Bonferroni correction (* p<0.05, ** p<0.01, *** p<0.001).

and average number of unique words per response (t=8.91, p<0.05), where *Case* individuals receive 11% more words, and 9% more unique words per response. This supports our hypothesis on the differences in verbosity, suggesting longer responses and lower repeatability of words are more likely to help psychosocial improvement.

Immediacy We find no significant difference in immediacy or the time to respond to posts. This could be because we study long-term and averaged-out improvements of psychosocial outcomes, rather than short-term bursts. Again, platform-specific design and post ranking on homepage plausibly does not distinguish the type and criticality of concern leading to all posts being responded back in similar intervals of time. This phenomenon is further revealed by the low standard deviation (∼12 mins.) in the time-to-first-response across all the responses in our TalkLife dataset. That said, immediacy is considered to be essential for coping with critical and crisis circumstances, and this leaves room for future investigations on the prevalence of such instances on TalkLife.

Diversity/ Creativity Supporting our next hypothesis, we find that the responses received by *Case* individuals are typically more diverse. The average distance (or diversity) among the responses received by the *Case* individuals is 5% higher (t=9.18, p <0.05). This also hints at the possibility that the *Control* individuals received more generic and templated responses. To understand this better in context, we inspect a few top keywords in responses received by the *Case* and *Control* users, to find many generic responses such as *hope great day*, *wish good luck*, etc. in responses to *Control* individuals.

Emotionality As Table 3 indicates, we find significant differences in the expression of positive affect — *Case* individuals received 3.5% greater positive affect. This aligns with literature that greater positivity is associated with effective

psychotherapy (Truax and Carkhuff 2007). In contrast, we find no significant differences in emotionality across anger, anxiety, sadness, negative affect, and swear. Nonetheless, a common trend across all the affective attributes is that the responses received by the *Case* individuals show a greater occurrence than that by *Control* individuals. Together, our hypothesis on emotionality is only partially supported.

Adaptability We measure adaptability of the responses in terms of topical congruence and linguistic style accommodation. Topical congruence occurs 6.6% higher in the responses received by the *Case* individuals than the *Control* individuals (t=11.55, p<0.05). In terms of linguistic accommodation, the responses received by the *Case* individuals show 31.15% greater (t=18.14, p<0.05) linguistic style matching than the ones received by the *Control* individuals. Both the measures of adaptability, therefore, support our hypothesis, aligning with prior work that greater adaptability in responses is associated with increased supportive outcomes.

Credibility of the Responders To examine if responder credibility significantly varied between *Case* and *Control* responses, we measure the differences in the responders' tenure (number of days on the platform), interactivity, number of posts, and the frequency of posting behavior (posts per day). Among these, we find no significant differences in the tenure and the number of posts per day. However, we find 16% greater interactivity and 10% greater number of posts for the responders to *Case* individuals as compared to that to the *Control* individuals. This suggests that responses from members who are more active on the platform seem to be typically more effective. Drawing on prior work, it may be associated with the fact that the members who are more experienced with the platform use more linguistically accommodating language or probably learn over time in what constitutes more supportive responses. Supporting our hypothesis, we find that *Case* individuals greatly received responses from those who are experienced "care and support" givers in the community.

Language Style of Responses We find that the average Categorical Dynamic Index (`CDI`) of responses received by *Case* individuals is 17% lower. This suggests that the responses received by *Case* are more dynamic in nature, or exhibit a dynamic style of thinking including a focus on others (such as greater use of pronouns), time-based stories, and use of simpler words (Pennebaker et al. 2014). Supporting our hypothesis, we conjecture that dynamic style of writing is likely to help psychosocial improvements on the platform.

Nature of Support We find that responses to *Case* is higher in both emotional and informational support. Among these, *Case* individuals receive 18% greater emotional support, and 25% greater informational support. Therefore, our hypothesis is supported, and we conjecture that greater support contributes to better psychosocial improvement on the platform. While prior work (De Choudhury and Kıcıman 2017) compared the efficacy of emotional against informational support, our work finds that both kinds of support are effective towards psychosocial improvement.

Summary We find that many of the treatment measures positively impact long-term psychosocial outcomes. We can use Cohen's d to rank treatments by their efficacy. Based on *Case* and *Control* means from Table 3, we can construct a binary treatment with a mean-split threshold on *Case* and

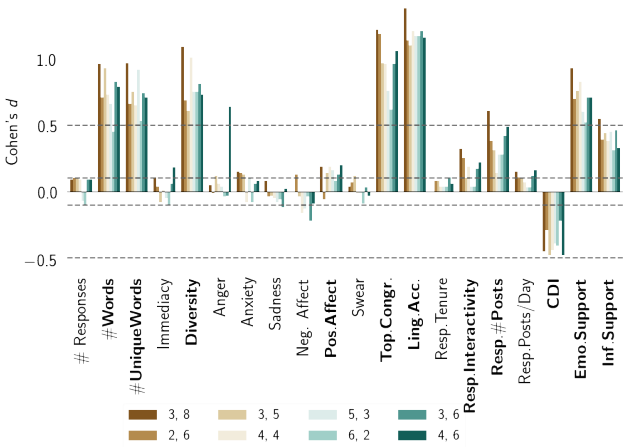


Figure 7: Cohen's d of treatment differences in responses received by *Case* and *Control* users for several combinations of (n_1, n_2) breakpoints defining pre- and post- treatment phases. Boldfaced measures show statistical significance in t-test and KS-test ($p < 0.05$). Statistically significant measures show consistent directionality in differences.

Control means. This maybe interpreted as treatments with high Cohen's d will likely ensure higher fraction of (binary) treated individuals with outcomes similar to the *Case* group. Our results indicate adaptability, diversity, verbosity, and emotional support rank highest in differentiating *Case* and *Control* individuals and thus can be considered as preferable treatment candidates for a future randomized experiment.

Robustness of Findings

Recall that our study design relies on chosen values of n_1 and n_2 posts to define pre-treatment, treatment, and post-treatment phases (Figure 2). We test if our findings hold robust for a variety of (n_1, n_2) combination pairs. For different pairs, we re-conduct our entire study including measuring outcomes, conducting matching, testing balance, and computing differences in treatment for matched *Case* and *Control*. Figure 7 shows the differences as effect-size (Cohen's d) across (n_1,n_2) pairs of (3,8), (2,6), (3,5), (4,4), (5,3), (6,2), (3,6), and (4,6). We find that effect size is very similar across all combinations of (n_1, n_2), showing a low standard deviation of 0.08 on average. Again, the treatment measures consistently show similar statistical significance as per t-test and KS-test. All these significant measures also show the *same* directionality of differences, such as verbosity, diversity, and adaptability are uniformly greater, and `CDI` is uniformly smaller for responses received by *Case* individuals.

Another component of our work concerns the decision to separately estimate the treatment differences between Case and Control for each treatment, rather than considering their effects together and also including any interaction effects. While we measure the differences for our theory-driven treatments independently, these treatment measures can be correlated and interdependent, e.g., positive affect and emotional support. Our study design is motivated towards providing interpretative understanding of how different psychotherapeutic measures function towards psychosocial improvement in OMHCs. Still, to test the robustness of such independent comparisons, we construct regression models with treatment

Measure	Coefficient	Measure	Coefficient
Num. Words	$7.32e-5^{***}$	Topical Congruence	$2.11e-2^{***}$
Diversity	$1.03e-6^{**}$	Resp. Interactivity	$2.01e-5^{*}$
Emotional Support	$5.55e-3^{***}$	CDI	$-3.52e-5^{**}$
Informational Support	$8.33e-4^{*}$	Anger	$-4.72e-3^{*}$

Table 4: Coefficients of linear regression of treatment measures as independent variables and psychosocial outcomes as dependent variables. Only statistically significant coefficients are reported (* $p<0.05$, ** $p<.01$, *** $p<0.001$).

measures as independent variables and overall psychosocial outcome as (continuous) dependent variable. We control our models with the same covariates used in matching. We eliminate correlated features using variance inflation factor (threshold=10) (Das Swain et al. 2019, Miles 2014). We also include regularizations (L1, L2) and interaction terms (degree 2) in the regression models. We find that all the interaction terms show statistically insignificant effects. The regularized and unregularized models show similar coefficients. For linear interpretability, Table 4 reports the unregularized model's coefficients of treatment measures found to be significant. The magnitude of regression coefficients is interesting and inspires further theoretical and empirical investigations. Consistent with our previous analysis, we find that the directionality of the regression coefficients agree with that we found by independently testing the treatment measures (Table 3).

The consistency of results via different approaches confirm that our findings are robust and not sensitive to choice of treatment periods or specific estimation methods, but rather a reflection of the phenomenon in our context of TalkLife.

7 Discussion and Conclusion

We studied factors that contribute to psychosocial changes in online mental health communities (OMHCs) using a case-control design. We examined whether effective support factors identified in psychotherapy literature are also similarly effective in OMHCs. Confirming past work, we find that factors such as diversity, adaptability, positivity, supportive nature, and dynamicity of language in responses are positively associated with effective psychosocial support. In contrast, simple factors such as immediacy and quantity of responses show insignificant effects on psychosocial outcomes. Our findings can be used to rank potential interventions for peer supporters. We discuss these points below.

Methodological implications. Our work provides a useful alternative to cohort-based analyses for studying cause-and-effect in online communities, especially when the outcome is well-specified and treatments are continuous variables. In such cases, our approach can help study two dimensional changes in the treatments — 1) along the breadth, that is studying several treatments together, and 2) along the depth, that is how much of a treatment is necessary.

From a treatment dosage perspective, each measure considered in our study is a continuous variable, and it is often not possible to determine an appropriate binary cutoff for (no) treatment in a prospective causal-inference setup. Our approach avoids this limitation by focusing on the differences in the treatment measures in *Case* and *Control* groups — and suggesting the dosage of measure required for desirable outcomes across a subpopulation (Table 3). This can be useful for design interventions on a social computing platform. For instance, TalkLife can propose guidelines that recommend expectations to the members in what ways they would be helped. Also, such differences can be used to formulate treatment cutoff in conducting careful experimental studies to verify and adopt design changes. More generally, our approach allows examining several treatments that potentially contribute to the same desirable outcome. Because the effects of each treatment can vary across individuals, such a study design helps to identify which treatment or combinations of treatments could be effective for certain individuals.

Implications for OMHCs. Implications for OMHCs. Given that OMHCs largely rely on amateur peer supporters, one of the biggest questions is how to help supporters write more effective responses. By comparing factors that lead to a positive outcome, our work provides evidence on effective support factors. We provide a way to rank and compare potential interventions so that effective candidate treatments can be considered and encouraged. For example, based on our results, OMHCs may nudge members to write more positive or adaptive responses. Our work also contributes to digital therapeutics, given limited availability of trained psychotherapy providers, we believe insights drawn from our work can be useful to train peer-supporters and volunteers who want to help in OMHCs (Kazdin 2011; Torous and Hsin 2018). That said, causal evidence from observational studies comes with the assumption that all confounders were conditioned. As randomized experiments are the gold standard to measure efficacy, our work provides a means to prioritize which treatments to consider for such experiments of understanding effective interventions.

Towards personalized support. From the perspective of individualized and precision medicine, our work builds the case for patient-centered and personalized psychotherapeutic care (Shippee et al. 2012). We find that, just like in face-to-face settings, templated and generic responses are not as effective as personalized and adaptive responses. Applying our approach of stratifying (or clustering) individuals based on psycholinguistic and psychosocial similarity may enable decision-making on what combination of treatments can be more effective in particular clusters. This can help design frameworks to tailor treatments per cluster of individuals.

Ethical Implications. Despite the potential, there are important ethical implications associated with using such quantitative analyses in practice. Privacy considerations should be made when machine guided interventions are tailored to OMHC participation. We expect analyses to be over de-identified datasets and interventions to be restricted to the online platform, ensuring that such analyses cannot be used to monitor one's trajectory of mental health and make offline decisions based on it. There are potential civil and ethical liability concerns in providing machine-guided support in an online medium, leaving room for further discussions on adopting these approaches in practice (Chancellor et al. 2019).

Limitations and Future Work. Our work has limitations, which also suggest promising future directions. We do not account for spill-over and passive engagement effects, e.g., individuals may be helped by browsing discussion threads of mental health support. We only consider mean-aggregated psychosocial outcomes, which is likely unable to capture shorter changes of psychosocial outcomes, e.g., individuals who show show fluctuating affective states or mood instability, which maybe accounted for by using complementary

data sources (Morshed et al. 2019). Our operationalization does not unpack intricacies in each psychosocial outcome separately, and does not encompass all mental health conditions; certain psychosocial changes (e.g., activity) considered to be positive in our study may not be applicable in certain mental health conditions (e.g., ADHD). Future work can address these concerns by examining psychosocial outcomes per condition, in a fine-grained temporal fashion.

Because our work examines observable behavior on TalkLife and plausibly excludes offline and latent individual differences, we cannot establish clinical validity. Future work obtain consented data (Saha et al. 2019) and expert-appraisal (Ernala et al. 2018, Levonian et al. 2020) to validate our findings with greater rigor. We only study those who show continued participation on the platform. This "dropouts" issue is also encountered in experimental and randomized trial settings (Lindsey 2000), and future work can include measures like likelihood of failed interactions (Zhang et al. 2018) and survival analysis to incorporate the behavior and dropping out of participants from platform (Ma et al. 2017, Yang et al. 2017). For treatments, we can examine the effects of social ties, social capital (Burke et al. 2010), cross-cultural accommodation (Pendse et al. 2019), and other linguistic attributes, such as politeness (Zhang et al. 2018), stance (Pavalanathan et al. 2017), and brevity (Gligorić et al. 2019).

As any other observational study, we recognize that we do not infer "true causality". We cannot eliminate the likelihood of type II errors, a vulnerability of retrospective causal design. Gelman and Rubins (2013) argue that reverse-causal problems are better studied with forward-causal questions, and Watts (2014) notes the impossibility to test all explanations simultaneously. While acknowledging these concerns, we believe our work is a step towards understanding the effects of a variety of heterogeneous factors in psychosocial outcomes. Accounting for all possible confounds is technically infeasible, and our work only *minimizes* the confounds by using a variety of theory-driven covariates, thereby providing insights beyond simpler correlational analyses. Alternative study designs such as instrumental variable methods may help to further minimize confounding biases. While our study only considers a finite set of treatment measures, more measures can be easily plugged in to understand their effectiveness. Our study design facilitates a simple but robust mechanism to understand the factors associated with psychosocial outcomes in an online setting, and in turn helps us draw actionable insights and implications towards running confirmatory randomized experiments and designing effective OMHCs.

8 Acknowledgement

Saha conducted this work while at Microsoft. We thank TalkLife for their support. We thank Monojit Choudhury, Munmun De Choudhury, Daejin Choi, Sindhu Ernala, Emre Kıcıman, Vedant Das Swain, Taisa Kushner, Sachin Pendse, Ian Stewart, Adith Swaminathan, and Dong Whi Yoo for their help and feedback.

References

Althoff, T.; Clark, K.; and Leskovec, J. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *TACL*.

Andalibi, N.; Haimson, O. L.; De Choudhury, M.; and Forte, A. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proc. CHI*.

Bandura, A. 1993. Perceived self-efficacy in cognitive development and functioning. *Educational psychologist* 28(2):117–148.

Baumel, A.; Tinkelman, A.; Mathur, N.; and Kane, J. M. 2018. Digital peer-support platform (7cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. *JMIR mHealth and uHealth* 6(2):e38.

Breckler, S. J. 1984. Empirical validation of affect, behavior, and cognition as distinct components of attitude. *J. Pers. Soc. Psychol*.

Burke, M.; Marlow, C.; and Lento, T. 2010. Social network activity and social well-being. In *Proc. CHI*.

Cavanagh, K.; Belnap, B. H.; Rothenberger, S. D.; Abebe, K. Z.; and Rollman, B. L. 2018. My care manager, my computer therapy and me: the relationship triangle in computerized cognitive behavioural therapy. *Internet interventions*.

Chancellor, S.; Baumer, E. P.; and De Choudhury, M. 2019. Who is the "human" in human-centered machine learning: The case of predicting mental health from social media. *PACM HCI* (CSCW).

Chikersal, P.; Belgrave, D.; Doherty, G.; Enrique, A.; Palacios, J. E.; Richards, D.; and Thieme, A. 2020. Understanding client support strategies to improve clinical outcomes in an online mental health intervention. In *Proc. CHI*.

Das Swain, V., et al. 2019. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. *Proc. IMWUT*.

Das Swain, V.; Saha, K.; Reddy, M. D.; Rajvanshy, H.; Abowd, G. D.; and De Choudhury, M. 2020. Modeling organizational culture with workplace experiences shared on glassdoor. In *CHI*. ACM.

De Choudhury, M., and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*.

De Choudhury, M., and Kıcıman, E. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*.

De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G.; and Kumar, M. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *CHI*.

Dinakar, K.; Chen, J.; Lieberman, H.; Picard, R.; and Filbin, R. 2015. Mixed-initiative real-time topic modeling & visualization for crisis counseling. In *IUI*.

Dos Reis, V. L., and Culotta, A. 2015. Using matched samples to estimate the effects of exercise on mental health from twitter.

Ernala, S. K.; Rizvi, A. F.; Birnbaum, M. L.; Kane, J. M.; and De Choudhury, M. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *CSCW*.

Ernala, S. K.; Labetoulle, T.; Bane, F.; Birnbaum, M. L.; Rizvi, A. F.; Kane, J. M.; and De Choudhury, M. 2018. Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. In *ICWSM*.

Ferrara, K. 1991. Accommodation in therapy. *Contexts of accommodation: Developments in applied sociolinguistics*.

Flannery, R. B., and Everly, G. S. 2000. Crisis intervention: A review. *International Journal of Emergency Mental Health*.

Gelman, A., and Imbens, G. 2013. Why ask why? forward causal inference and reverse causal questions.

Glass, T. A., and Maddox, G. L. 1992. The quality and quantity of social support: stroke recovery as psycho-social transition. *SSM*.

Gligorić, K.; Anderson, A.; and West, R. 2019. Causal effects of brevity on style and success in social media. *PACM HCI* (CSCW).

Guntuku, S. C.; Buffone, A.; Jaidka, K.; Eichstaedt, J. C.; and Ungar, L. H. 2019. Understanding and measuring psychological stress using social media. In *Proc. ICWSM*.

Haberstroh, S.; Duffey, T.; Evans, M.; Gee, R.; and Trepal, H. 2007. The experience of online counseling. *JMHC*.

Hannan, E. L. 2008. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC*.

Hernan, M. A., and Robins, J. M. 2010. *Causal inference*.

Huh, J. 2015. Clinical questions in online health communities: the case of see your doctor threads. In *Proc. CSCW*.

Kazdin, A. E. 2011. Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist* 66(8):685.

Kiciman, E.; Counts, S.; and Gasser, M. 2018. Using longitudinal social media analysis to understand the effects of early college alcohol use. In *ICWSM*.

Kolden, G. G.; Klein, M. H.; Wang, C.-C.; and Austin, S. B. 2011. Congruence/genuineness. *Psychotherapy* 48(1):65.

Kummervold, P. E.; Gammon, D.; Bergvik, S.; Johnsen, J.-A. K.; Hasvold, T.; and Rosenvinge, J. H. 2002. Social support in a wired world: use of online mental health forums in norway.

Kushner, T., and Sharma, A. 2020. Bursts of activity: Temporal patterns of help-seeking and support in online mental health forums. In *WebConf*.

Labov, W., and Fanshel, D. 1977. *Therapeutic discourse: Psychotherapy as conversation*. Academic Press.

Lambert, M. J., and Barley, D. E. 2001. Research summary on the therapeutic relationship and psychotherapy outcome. *Psychother*.

Larson, J. S. 1996. The world health organization's definition of health: Social versus spiritual health. *Social Indicators Research*.

Levonian, Z.; Erikson, D. R.; Luo, W.; Narayanan, S.; Rubya, S.; Vachher, P.; Terveen, L.; and Yarosh, S. 2020. Bridging qualitative and quantitative methods for user modeling: Tracing cancer patient behavior in an online health community. In *ICWSM*.

Lindsey, J. 2000. Dropouts in longitudinal studies: definitions and models. *Journal of biopharmaceutical statistics* 10(4):503–525.

Love, B.; Crook, B.; Thompson, C. M.; Zaitchik, S.; Knapp, J.; LeFebvre, L.; Jones, B.; Donovan-Kicken, E.; Eargle, E.; and Rechis, R. 2012. Exploring psychosocial support online: a content analysis of messages in an adolescent and young adult cancer community.

Ma, H.; Smith, C. E.; He, L.; Narayanan, S.; Giaquinto, R. A.; Evans, R.; Hanson, L.; and Yarosh, S. 2017. Write for life: Persisting in online health communities through expressive writing and social support. *PACM HCI* (CSCW).

Ma, X.; Cheng, J.; Iyer, S.; and Naaman, M. 2019. When do people trust their social groups? In *CHI*.

Merolli, M.; Gray, K.; and Martin-Sanchez, F. 2013. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *J. Biomed. Inform.*

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*.

Miles, J. 2014. Tolerance and variance inflation factor. *Wiley StatsRef: Statistics Reference Online*.

Morshed, M. B.; Saha, K.; Li, R.; D'Mello, S. K.; De Choudhury, M.; Abowd, G. D.; and Plötz, T. 2019. Prediction of mood instability with passive sensing. *PACM IMWUT*.

Nambisan, P. 2011. Information seeking and social support in online health communities: impact on patients' perceived empathy. *JAMIA*.

Norcross, J. C., and Lambert, M. J. 2018. Psychotherapy relationships that work iii. *Psychotherapy* 55(4):303.

Oh, H. J.; Lauckner, C.; Boehmer, J.; Fewins-Bliss, R.; and Li, K. 2013. Facebooking for health: An examination into the solicitation and effects of health-related social support on social networking sites. *Comput. Hum. Behav*.

Olteanu, A.; Varol, O.; and Kiciman, E. 2017. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proc. CSCW*.

Pavalanathan, U.; Fitzpatrick, J.; Kiesling, S.; and Eisenstein, J. 2017. A multidimensional lexicon for interpersonal stancetaking. In *ACL*.

Pearl, J. 2009. Causal inference in statistics: An overview.

Pendse, S. R.; Niederhoffer, K.; and Sharma, A. 2019. Cross-Cultural Differences in the Use of Online Mental Health Support Forums. *PACM HCI* (CSCW).

Pennebaker, J. W.; Chung, C. K.; Frazee, J.; Lavergne, G. M.; and Beaver, D. I. 2014. When small words foretell academic success: The case of college admissions essays. *PloS one* 9(12):e115844.

Pennebaker, J. W.; Mehl, M. R.; and Niederhoffer, K. G. 2003. Psychological aspects of natural language use: Our words, our selves. *Annu. Rev. Psychol.*

Perkins, R. 2001. What constitutes success?: The relative priority of service users' and clinicians' views of mental health services. *Br. J. Psychiatry*.

Potts, H. W. 2005. Online support groups: an overlooked resource for patients. *He@ lth Information on the Internet*.

Pruksachatkun, Y.; Pendse, S. R.; and Sharma, A. 2019. Moments of change: Analyzing peer-based cognitive support in online mental health forums. In *CHI*.

Rollman, B. L.; Belnap, B. H.; Abebe, K. Z.; Spring, M. B.; Rotondi, A. J.; Rothenberger, S. D.; and Karp, J. F. 2018. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. *JAMA psychiatry*.

Rubin, D. B. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Am. Stat. Assoc*.

Saha, K., and De Choudhury, M. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *PACM HCI* (CSCW).

Saha, K., et al. 2019. Social media as a passive sensor in longitudinal studies of human behavior and wellbeing. In *CHI Ext. Abstracts*.

Saha, K.; Chan, L.; De Barbaro, K.; Abowd, G. D.; and De Choudhury, M. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proc. ACM IMWUT*.

Saha, K.; Kim, S. C.; Reddy, M. D.; Carter, A. J.; Sharma, E.; Haimson, O. L.; and De Choudhury, M. 2019a. The language of lgbtq+ minority stress experiences on social media. *PACM HCI* (CSCW).

Saha, K.; Sugar, B.; Torous, J.; Abrahao, B.; Kıcıman, E.; and De Choudhury, M. 2019b. A social media study on the effects of psychiatric medication use. In *ICWSM*.

Saha, K.; Ernala, S. K.; Dutta, S.; Sharma, E.; and De Choudhury, M. 2020. Understanding moderation in online mental health communities. In *HCII*. Springer.

Saha, K.; Weber, I.; and De Choudhury, M. 2018. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. In *ICWSM*.

Satopaa, V.; Albrecht, J.; Irwin, D.; and Raghavan, B. 2011. Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *ICDCS*.

Schulz, K. F., and Grimes, D. A. 2002. Case-control studies: research in reverse. *The Lancet*.

Sharma, E., and De Choudhury, M. 2018. Mental health support and its relationship to linguistic accommodation in online communities. In *CHI*.

Shippee, N. D.; Shah, N. D.; May, C. R.; Mair, F. S.; and Montori, V. M. 2012. Cumulative complexity: a functional, patient-centered model of patient complexity can improve research and practice. *J. Clin. Epidemiol.*

Silber, J. H.; Rosenbaum, P. R.; Trudeau, M. E.; Even-Shoshan, O.; Chen, W.; Zhang, X.; and Mosher, R. E. 2001. Multivariate matching and bias reduction in the surgical outcomes study. *Med. Care*.

Steinfield, C.; Ellison, N. B.; and Lampe, C. 2008. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *JADP*.

Tan, C.; Niculae, V.; Danescu-Niculescu-Mizil, C.; and Lee, L. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proc. WWW*.

Torous, J., and Hsin, H. 2018. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. *NPJ digital medicine*.

Truax, C. B., and Carkhuff, R. 2007. *Toward effective counseling and psychotherapy: Training and practice*. Transaction Publishers.

Watts, D. J. 2014. Common sense and sociological explanations. *American Journal of Sociology* 120(2):313–351.

Winzelberg, A. J.; Classen, C.; Alpers, G. W.; Roberts, H.; Koopman, C.; Adams, R. E.; Ernst, H.; Dev, P.; and Taylor, C. B. 2003. Evaluation of an internet support group for women with primary breast cancer. *Cancer*.

Yang, D.; Kraut, R.; and Levine, J. M. 2017. Commitment of newcomers and old-timers to online health support communities. In *Proc. CHI*.

Yoo, D. W., and De Choudhury, M. 2019. Designing dashboard for campus stakeholders to support college student mental health. In *Pervasive Health*.

Zhang, J.; Chang, J.; Danescu-Niculescu-Mizil, C.; Dixon, L.; Hua, Y.; Taraborelli, D.; and Thain, N. 2018. Conversations gone awry: Detecting early signs of conversational failure. In *Proc. ACL*.

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

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

Yozen Liu
Snap Inc.
Santa Monica, CA, USA
yliu2@snap.com

Nicholas Vincent
Northwestern University
Evanston, IL, USA
nickvincent@u.northwestern.edu

Farhan Asif Chowdhury
University of New Mexico
Albuquerque, NM, USA
fasifchowdhury@unm.edu

Leonardo Neves
Snap Inc.
Santa Monica, CA, USA
lneves@snap.com

Neil Shah
Snap Inc.
Santa Monica, CA, USA
nshah@snap.com

Maarten W. Bos
Snap Inc.
Santa Monica, CA, USA
mbos@snap.com

ABSTRACT

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

CCS CONCEPTS

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

KEYWORDS

social media, ads, Snapchat, momentary behaviors, causal-inference, matching, ad allocation

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

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445394

ACM Reference Format:
Koustuv Saha, Yozen Liu, Nicholas Vincent, Farhan Asif Chowdhury, Leonardo Neves, Neil Shah, and Maarten W. Bos. 2021. AdverTiming Matters: Examining User Ad Consumption for Effective Ad Allocations on Social Media. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3411764.3445394

1 INTRODUCTION

In recent years, many online platforms predominantly generate revenue from advertisements (ads). Ad revenue offsets costs, making the services "free" to use — ad-supported business models are even considered to be at the "heart of the free internet" [104]. Some common examples are search engine services like Google and Bing, and social platforms like Facebook, StackExchange, LinkedIn, and Snapchat. These online platforms show ads via a variety of implicit and explicit mechanisms, including "sponsored" or "promoted" content. However, an online platform that relies on ad revenue must contend with the tradeoff between ad revenue and ads' impact on users. If a platform shows more ads, it runs the risk of hurting the user experience and losing its userbase. [10, 12, 17]. Consequently, platforms resort to optimizing ad allocations that aim for multi-stakeholder benefits from user-centric, platform-centric, and advertiser-centric perspectives.

Traditionally, ad delivery in mass media such as print and television took a blanket approach — the same ad was shown to everyone who read the same newspaper or watched the same television channel at a time, and accordingly, demography (age/gender)-based targeting was conducted using people's interests (e.g. certain ads might only play on a sports channel). Although a blanket approach to advertising somewhat works in new and online media, online platforms introduce new complexities, potentials and dimensions [10]. With the ubiquity of various personalizations, content and ad delivery is often algorithmically customized to suit the interests of a specific user and improve their engagement with the platform. Importantly, the objective of personalized ad allocation is not just to

increase revenue per user (and ad), but also to improve a user's satisfaction and experience with ads — a user would plausibly be more interested in ads topically relevant to their interests [61]. Personalized ad allocation has therefore built on the success of research in Recommender Systems, Machine Learning, and Human-Computer-Interaction (HCI) [21, 36, 46]. However, content personalization algorithms typically rely on demographic attributes like age, race, and gender, which are not only privacy-intrusive but are also static and exclusionary. As such, this practice has been critiqued in the fairness, accountability, transparency literature as reinforcing (and potentially exacerbating) stereotypes and societal biases [3, 41, 82].

While demographic attributes and content-based recommendations have been tremendously explored, other factors remain relatively less known — online ad allocation and ad spacing strategies typically rely on sets of rules, such as ads are shown after T duration or after N content views, where the parameters are mostly drawn from observations about the average user on a platform. However, the same ad allocation strategy would not necessarily work effectively on all users, given that every individual is different, and that they have a different lifestyle, behavior, needs, and engagement both offline and online [6]. In fact, online behaviors are also functions of offline context and routines, as well as users' momentary psychological and cognitive states [37, 64, 91]. Therefore, it is important to embrace and evaluate dynamic and context- centric ad allocation strategies. This line of work remains underexplored both generally and particularly on evolving and newer forms of social and online content delivery.

Our work aims to address the above gap by studying ad consumption on Snapchat, a popular multimedia-driven online social platform which has a content feed called "Discover Feed", similar to content feeds offered by Facebook, Twitter, and Instagram's News or Story feeds. We latch on to the notion that users may show varying ad-consumption behaviors by the time of the day or other in-the-moment activities on the platform. Theoretically, our work is motivated by the body of literature that explains the psychological and time of the day effects in content and ad consumption [36, 102]. Practically, our work builds on the motivation that by teasing out these effects in a person-centric and context-centric fashion, we can not only draw better insights about ad consumption but can also make specific recommendations about *when* and *whom* questions in "What be shown to whom, and when?" — a question that interests both content providers (platform owners) as well as consumers (users). In particular, our work proposes three *research aims*:

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

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

Methodologically, this work contributes a novel causal-framework to infer and model ad consumption by minimizing the confounding factors. This approach can be extended to study other forms of user engagement using observational social media data. Additionally, we propose a computational approach to obtain preferred times of ad consumption for different users. Our results reveal the potential to use our method in ad engagement and other engagement mechanisms. Our work provides insights into ad consumption on a social platform, as well as makes recommendations for effective ad allocation using on-platform user behavior.

From an HCI perspective, our work augments the body of work studying the need to balance user experience and ad effectiveness [4, 56, 88, 106]. Theoretically, our work bears implications in advancing the understanding of ad consumption on social media. We discuss the implications to design better and more responsible ad allocation systems from multi-stakeholder standpoints. By taking a broader view of ad allocation, we argue that it is possible to create better outcomes for both users and platforms.

Applied to social platforms, our ad allocation approach's efficacy would help to accomplish monetary goals with fewer ads, and therefore, can lead to allocating fewer ads —– a solution preferred by users. Better ad interaction would potentially lead to better social interaction and user experience on the online social platforms. Finally, a key advantage of our computational framework lies in the fact that it does not use demographic, typographic, or privacy-sensitive information of the users. We also discuss the privacy-preserving component and ethical implications of this work.

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

2 BACKGROUND AND RELATED WORK

2.1 Ad Effectiveness and User Experience

Conceptually, ad effectiveness is a key indicator of success of an ad based on how well it does or the returns it generates in various possible outcomes such as user likeability, engagement, and sales [14, 105]. In an early work, Morrison and Dainoff studied ad complexity

and dwell times, i.e., how much time a user spends to look at an ad and if they remember an ad more than others [72]; dwell times have been widely used as an implicit metric to study user interest and satisfaction [45, 55]. Doyle and Saunders defined effective ads as those that help advertisers reach their goals [22]. Ducoffe developed survey scales to measure ad effectiveness in terms of ad value in traditional media [23], which was later extended in the online media [24], positing ad value as a form of *communication engagement* between advertisers and consumers [24, 25]. Ducoffe and Curlo followed up to propose quantifiable concepts of expected advertising value (EAV) and outcome advertising value (OAV) of ads [25]. These assessments have also been used in practice and in comparing online and offline ad effectiveness. In the online form, ad effectiveness is often quantified as return rate or Click-Through-Rate (CTR) [87]. CTR measures the proportion of effectively allocated ads or the ratio of the clicks on an ad to its number of impressions [15]. Other work has proposed sophisticated measurements of online ad effectiveness such as using ghost ads and experimental approaches [43].

Ad effectiveness is considered a vital outcome while planning, creating, and executing an ad [83]. Research has studied how various factors relate with ad value [25, 63, 84]. These studies found that informativeness and entertainment aspects bear a positive association with ad effectiveness, whereas intrusiveness bears a negative relationship [24]. Further, ad effectiveness shares a deep and complex interplay with user experience on the platform [10, 73]. Brajnik and Gabrielli reviewed the effects of online advertising on user experience and proposed a systematic theoretical framework for its better understanding [10]. Ads can cause fatigue, irritation, and negative emotions for users, making them leave and reduce engagement on the platform, consequently hurting both ad effectiveness and platform engagement [10]. Therefore, it is critical to optimize ad allocation in such a way that user experience is not compromised, as shown in recent HCI research through gamification [4], intelligent placement strategies [73], and animation [21].

The above body of work motivates us to operationalize and study factors associated with ad effectiveness on social media. We extend the HCI community's long-drawn interest in balancing user experience with ad effectiveness [4, 17, 21, 73, 88, 106]. We define ad effectiveness using two measures based on 1) what fraction of time a user fully watches an ad (or ad reception), and 2) whether a user expresses some form of interest in the ad by clicking on it (or ad CTR). We then examine the role of (previously unexplored) factors such as timing and momentary on-platform behavior in explaining ad effectiveness outcomes in online platforms.

2.2 Ad Consumption Contextualized with Psychological Factors

Marketing and consumer research has extensively studied the importance of "antecedent state" — a term that encompasses all of the momentary financial, psychological, and physiological attributes with which a consumer arrives at a marketing interaction [8]. Haugtvedt et al. studied how personality traits associate with ad effectiveness [39]. In particular, mood states are known to significantly influence consumer behavior, judgment, and recall [29], and within the space of online ads, beliefs and attitude towards ads have been identified to predict ad effectiveness [26, 111]. Batra and Stayman showed that positive mood mediates brand attitudes in print ads [7], and Edwards et al. adopted the lens of psychological reactance to understand forced responses to ads and correspondingly the perceived intrusiveness and irritation to ads [11, 27]. People's responses to ads include affective, behavioral, and cognitive components [24, 98, 111]. Here, the affective component includes irritation and entertainment elicited by an ad [27], the behavioral component includes pre- and post- ad purchasing behavior [98], and the cognitive component includes factors like informativeness of an ad [24]. Relatedly, Bronner et al. studied the relationship between mood and ad effectiveness [13].

Parallelly, a body of research notes how time of day may affect the variety-seeking behavior of individuals [35]. In fact, circadian orientation and time of day are known to associate with an individual's depth of information processing with respect to ads [16]. Prior research studied how ad effectiveness varies with time of day by different age groups of individuals [33], and Tellis et al. studied the microeffects of time, content, and duration on ad effectiveness [102]. Relatedly, Kapoor et al. noted the promises of just-in-time recommendations in online platforms [46, 47]. Taken together, these studies explain how ad consumption is dependent on several contextual and psychological factors.

While the role of context in explaining ad effectiveness has been extensively studied in offline and traditional forms of media, it still remains an underexplored avenue in the space of social media and online platforms. In fact, with the emergence of newer forms of media and content delivery, it is important to assess contextual factors and accordingly improve content delivery to ensure better user experience [68]. Further, due to the lack of a comprehensive understanding of how users consume ads on these new online content delivery platforms coupled with ubiquitous technological affordances (such as smartphones and wearables), ad allocation on social media is still largely driven by only static rules and content-related personalization. Our work aims to address this gap in theory and practice by examining ad consumption with respect to time and momentary factors on Snapchat. We further simulate an experiment that evaluates the efficacy of our context-centric factors in making effective ad allocations.

2.3 Social Media and Observational Data

Research reveals how social media activities reflect people's offline routines and behaviors [18, 37, 64]. Social media behaviors can potentially reveal naturalistic patterns of behavior, cognition, psychological states and social milieu, both in real-time and across longitudinal time [32, 59, 92]. Prior work has harnessed social media to infer individual-centric attributes ranging across personality traits and wellbeing using machine learning and computational linguistics [34, 80, 81, 96, 99]. Kosinski et al. used Facebook Likes to predict a range of attributes such as sexual orientation, ethnicity, personality, intelligence, addictive behavior, age, and gender.

In the related problem space as ours, social media behaviors can explain ad consumption and vice versa [21, 100, 110, 112]. Kim et al. investigated the antecedents of clicking ads on Facebook [50] and Mao and Zhang studied the factors associated with users' intention to click on social media ads, particularly around content,

media, and individual-related factors [62]. Prior work has also examined social media ads with respect to perceived informativeness, entertainment, and irritation [50, 63]. and Youn and Kim examined reactance related factors of avoiding ads on Facebook [112].

In general, the effect of an (external or internal) change or an intervention is measured using causal-inference approaches. These approaches draw motivation from epidemiological research settings of randomized controlled trials (RCTs): participants are randomly assigned to a treatment and a control group where the former receives a drug, and the latter receives a placebo, and then changes are measured in the two groups to quantify the effect of the drug [38]. Similarly, understanding user behavior on a platform due to platform-based interventions are best studied with experimental and A/B test approaches [57, 71]. However, such approaches bear caveats. For instance, experimental studies that seek participant consent can be limited by concerns of observer effect [1] — participants may modulate their otherwise normal behavior with the awareness being monitored or observed. Alternatively, experimental research conducted without participants' awareness are deemed unethical especially in the human-centered research paradigm [44, 69]. For example, the Facebook emotion contagion study did not inform the participants that their feeds would be modified for research [54]. While this work was successful in uncovering valuable insights regarding people's affective behavior on social media, this work was heavily critiqued on ethical grounds [44]. Further, experimentation without apriori awareness of impact on participants may lead to long-term negative consequences for both platforms and individuals.

Consequently, in problem settings where experimental approaches may be infeasible or unethical, observational studies can be an alternative. While observational studies cannot guarantee "true causality", they are designed in a way to minimize confounds and to investigate longitudinal data in providing stronger evidences than naive correlational analyses [42]. These studies can also benefit future randomized experiments where no preferred treatment is known apriori [89]. Recently, this kind of study has also generated interested in HCI, social, and behavioral science, including that using social media data [20, 48, 76, 90, 95, 114]. Notably, De Choudhury et al. examined the shifts in suicidal ideation tendencies in online communities [20] and Culotta estimated county health statistics using Twitter data [18, 20]. Of particular interest is Saha et al.'s work which motivates us to operationalize social media behavioral measures such as activity, interactivity, and interaction diversity, whose relationship we examine with ad reception in our study [96].

Our work draws motivation from the success of observational data and quasi-experimental study design to understand ad consumption on social media. Besides, we also note that our study values the importance of a contextualized person-centric design [93] which is not only an improvement over one-for-all or generalized approaches but also stays clear of using demographic and trait-based information of users. While using such information may though improve prediction accuracy in some problem settings, these approaches could be exclusionary, discriminating, privacy-intrusive, and unethical [41, 82]. Rather, our study design incorporates dynamic platform behaviors to draw insights corresponding to user strata exhibiting similar combinations of platform behaviors.

3 DATA

We conduct our study on the Snapchat platform. Snapchat is an online social and instant messaging platform that enables users to share and interact with others using ephemeral content, including text, images, videos, and other forms of multimedia. Snapchat is particularly very popular among the youth, with 73% of the 18-25 age demographic in the U.S. being Snapchat users [78]. Snapchat provides a Discover Feed where users can find and view recommended content in tiled story format from news publishers, brands, and content providers, such as ESPN, Wall Street Journal, Daily Mail, etc. Users can browse through these tiles, and when on a tile, they can consume, skip, or advance to the next recommended content. Snapchat's Discover Feed is design-wise similar to content-feed of Facebook, Twitter, or Instagram [2]. Discover Feed also shows ads which contribute to ∼98% of Snapchat's revenue [86]. As on most other platforms, users can watch an ad on Snapchat as long as they are interested, skip if they are uninterested or want to move on to other content, and/or swipe up (considered an ad click) if they are particularly interested to know more.

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

3.1 Preliminary Analysis

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

To visually examine the above variability in ad reception by hour and user, we cluster users on aggregated behaviors on the platform (such as the number of app opens, frequency, and amount of posting and consuming content on Snapchat). We adopt k-means clustering (k=`20`) where the number of clusters is roughly determined using the Elbow heuristic [97] (Figure 2b). Figure 2c plots a heatmap of the mean z-score of ad reception by hour, with each cluster of users on the vertical axis. We find that the bottom-most row in the heatmap or the ad reception at an overall level barely shows any variance across hours. However, the same does not hold true if we look at the rest of the rows with shades of light and dark distributed throughout. This suggests that ad reception shows different patterns both between and within clusters across hours.

These preliminary analyses motivate us to investigate if users have different "preferred times" of ad consumption (or times of

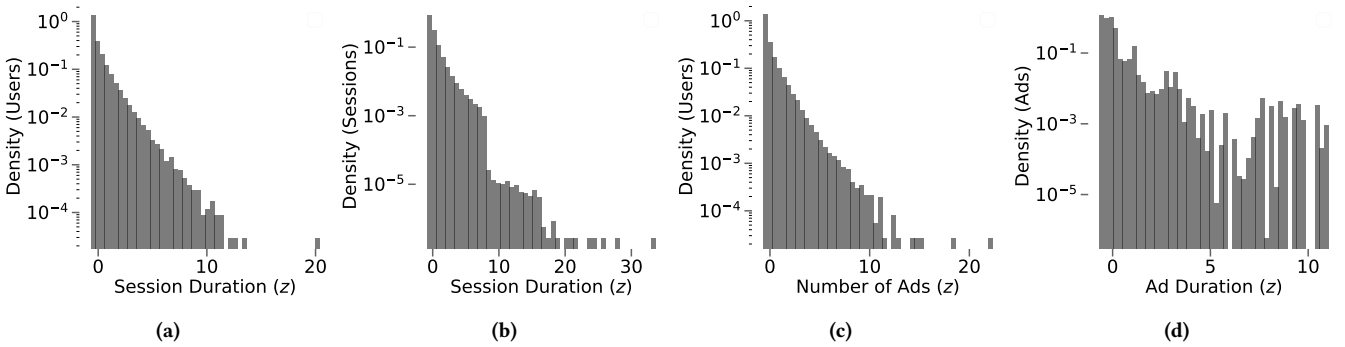


Figure 1: Distribution of data by z-scores of (a) Total duration on platform per user, (b) Duration per session, (c) Number of ads per user, (d) Ad duration per ad.

the day when ads might be less disruptive), and that clustering (or stratifying) users with on-platform attributes can provide key insights on ad consumption. These attributes do not use a user's demographic and trait-based information, and therefore, can be argued to be more privacy-preserving than traditional forms of user-profiling [77]. Concretely, any interventions using these insights would not access a user's personal data.

4 AIM 1: AD TIMINGS AND EFFECTIVENESS

4.1 Study Design and Rationale

Ads can disrupt a user's normal course of action on a social platform [4, 88]. Prior work has explored methodologies to improve ad effectiveness by showing personalized advertisements to individuals, where major approaches have largely focused on user interests and content-based personalizations (see Section 2). In this regard, context- and time- driven factors have remain largely unexplored, particularly on social media. Motivated by the role of context and time of day effects [16, 33, 102] and initial insights from our preliminary analysis, we hypothesize that different users have different preferred times of ad consumption on the platform.

Ideally, such a problem would be best examined in an experimental or A/B test setting; however, these methods have caveats [38]. For example, experimental allocations of ads may lead to unintended consequences such as changing platform experiences and risks of long-term perceptions about the platform. Again, these approaches are sensitive to particular parameters and thresholds, such as what quantity of ads can be shown and when — which remain unknown apriori to experimentation. Given these considerations, we draw on quasi-experimental approaches on observational data to understand the effect of ad allocations with respect to time preferences of users [89]. In particular, we adopt a causal framework based on matching, which simulates an experimental setting by controlling for as many covariates as possible [42]. This approach builds on the potential outcomes framework, examining if an outcome is caused by a treatment T by comparing two potential outcomes: (1) Y_i when exposed to T ($T = 1$), and (2) Y_i if there was no T ($T = 0$). Because it is impossible to obtain both kinds of outcomes for the same user, this framework overcomes this challenge by estimating the missing counterfactual for a user based on the outcomes of a matched user — a user with similar distribution of covariates but differing treatment status. We employ stratified propensity score analysis [76, 95] to match users and examine ad outcomes in Treated and Control groups of individuals. This section describes the methodological considerations and approach in detail.

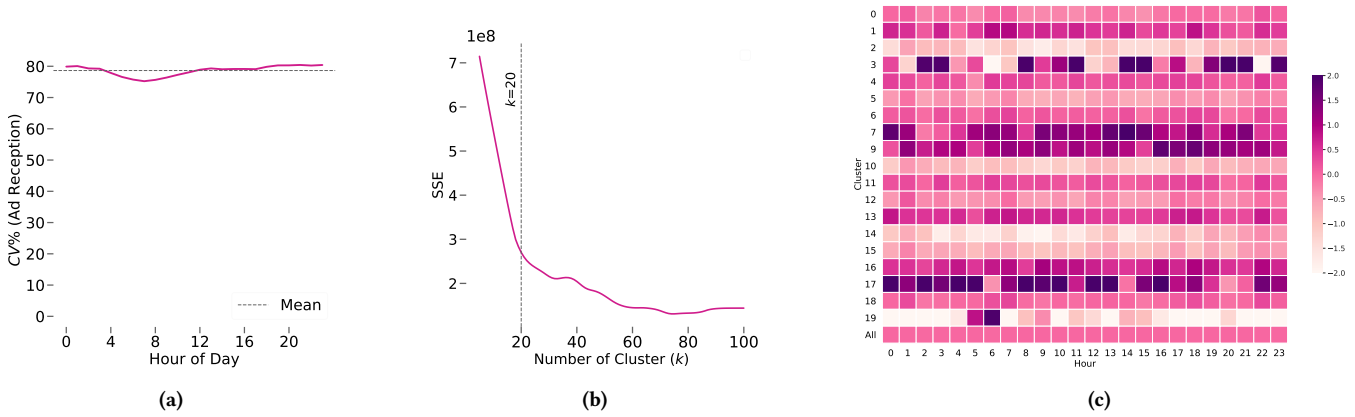
This paper communicates our insights using z-score-transformed quantities from raw data metrics due to privacy and sensitivity reasons. Importantly, z-scores are not sensitive to inconsistent magnitudes of absolute values, making normalized comparisons across multiple measures feasible. By definition, z-scores quantify the number of standard deviations by which the value of a raw score is above or below the mean. Similar standardization techniques have been adopted in prior social media studies [32]. z-scores are calculated as $(x-\mu)/\sigma$, where x is the raw value, μ and σ are respectively the mean and standard deviations of the population. Here, we obtain population μ and σ on the entire data per measure. We interpret positive z-scores as values above the mean, and negative z-scores as those below the mean.

4.2 Defining Outcomes: Ad Effectiveness

A causal study typically measures the *effect-of-a-cause*, and effects are measured as changes in *outcomes*. Our work measures outcomes in terms of *ad effectiveness*. We draw motivation from prior research that ad effectiveness is a function of how interested people feel in watching an ad, and what actions they take following their consumption of the ad (such as buying the product, or other behavioral markers indicating their interest in the product) [72, 105]. On the basis of this, we operationalize ad effectiveness using two measures — 1) *Ad Reception* or the proportion of time ads are watched over the total duration of ads in a session and 2) *Ad Click Through Rate (CTR)* as the proportion of ads that were clicked on (or that users swiped up on, in the case of Snapchat).

4.3 Defining Baseline & Measurement Periods

We aim to measure ad effectiveness by conditioning on how ads were allocated to users. For this purpose, we draw upon recent causal inference research on observational social media data [94], to define a Baseline and a Measurement period in the longitudinal timeline of each user (schematically represented in Figure 3). In the Baseline period, we aim to compute how users consumed ads shown at different hours of day (and weekdays). This allows us to obtain the preferred hours of ad consumption for each user. Then, in the Measurement period we measure the effect of showing ads in



(a) (b) (c)

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

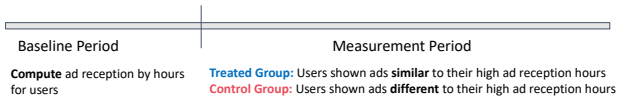
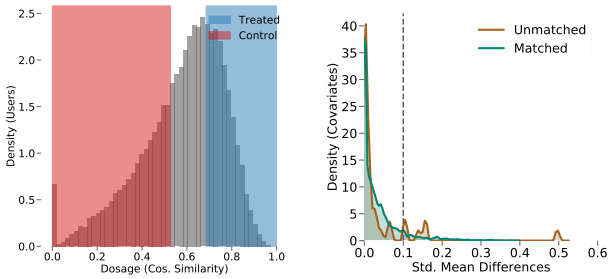


Figure 3: A schematic representation of splitting user longitudinal timelines into baseline and measurement periods.

preferred hours by minimizing the confounds due to ad quantity and user behavior on the platform. We choose to split the longitudinal timeline of 78Kusers before and after January 10, 2020, leaving us with roughly three weeks of data in the Baseline period and six weeks of data in the Measurement period for each user.

4.4 Defining Treatment: Treated & Control

As we examine the effect of timing ad allocations, our study design adopts a *Treatment Dosage* on the basis of preferred times of ad consumption. We operationalize the treatment dosage on how similarly (or differently) ads were allocated in the Measurement period with respect to a user's high (or low) hourly ad consumption in the Baseline period. This builds on the notion that a user who consumed ads well at H_1 hours and poorly at H_2 hours during the Baseline period, would show similar consumption patterns when ads are allocated to them at H_1 and H_2 hours in the Measurement period. For each "hour of the day" in each "day of the week" (henceforth, referred to as hour-weekday pair), we compute an aggregated average of ad quantity (number of ads normalized by number of browsed content tiles) and ad reception per user separately in the Baseline and Measurement periods. First, we obtain the hour-weekday wise vector of ad reception for each user in the Baseline period (v_1). Then, we obtain the hour-weekday wise vector of ad allocation for each user in the Measurement period (v_2). Finally, we define treatment dosage as the cosine similarity of vectors v_1 and v_2, computed per user. Essentially, the greater the dosage, the greater is the likelihood of ad allocation (in Measurement period) during a user's preferred ad reception hours (as inferred from ad consumption behavior in the Baseline period).



(a) Dosage (b) Matching Balance

Figure 4: Distribution of: (a) Dosage: cosine similarity of ad reception in Baseline period and ad allocation in Measurement; (b): Balance of covariates in matching.

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

4.5 Matching for Causal Inference

4.5.1 Matching Covariates. Matching aims to control for *covariates* so that the effects of treatment are examined between two comparable groups of users [42]. We note that ad outcomes can be confounded by factors such as how long someone stays in a session, or how they engage on Snapchat, or even how many ads were allocated. To mitigate such confounds in our analyses, we adopt an

approach called matching — when conditioned on high dimensional covariate data, matching can minimize biases compared to naive correlational analyses [42]. Our approach controls for a variety of covariates so that the compared Treated and Control show similar baseline behaviors. Drawing on prior work [48, 76, 96], we use 41 covariates (listed in the Appendix Table A1). The first set of covariates are based on aggregated Baseline data spanning across the number and frequency of app opens, social interactions, different types of interactions, etc. We also include a second set of covariates obtained from the Measurement data based on the number of sessions, average duration of sessions, and the average number of ads exposed, to ensure that matched Treated and Control users had comparable engagement on the platform and were exposed to similar quantity of ads.

4.5.2 Stratified Propensity Score Analysis. We use matching to find pairs (generalizable to groups) of Treated and Control users whose covariates are statistically very similar to one another. The propensity score model matches users based on their *likelihood* of receiving the treatment, or the propensity scores. Stratified matching potentially overcomes the challenges of exact one-to-one pair matching which can lead to biases [51]. Our stratified matching approach groups users with similar propensity scores into strata [48]. Every stratum, therefore, consist of users with comparable covariates. This approach allows us to isolate and estimate the effects of the treatment within each stratum.

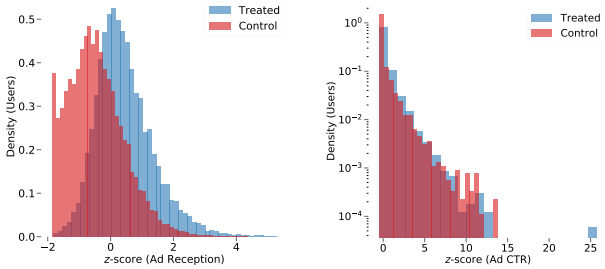
To compute the propensity scores, we build a logistic regression model that predicts a user's binarized treatment status (0 for Control and 1 for Treated) based on their covariates. We segregate the remaining distribution into 100 strata of equal width — and discard those strata containing less than 50 users which further ensures that our causal analysis per stratum remains restricted to a sufficient number of similar users, and therefore is minimally biased [95]. This leads us to a final matched dataset of 92 strata consisting of 16,003 Treated and 15,682 Control users in total.

4.5.3 Quality of Matching. To test that our matching yields statistically comparable Treated and Control users, we evaluate the balance of covariates. For each covariate, we compute the standardized mean difference (SMD) in the Treated and Control groups in each of the 92 valid strata. SMD calculates the difference in the mean covariate values between the two groups as a fraction of the pooled standard deviation of the two groups. Two groups are considered to be balanced if all the covariates reveal SMD lower than 0.2 [48, 101]. This condition is satisfied by a majority of the covariates in our matched datasets, and we obtain a 18.9% reduction in the SMDs of matched from unmatched samples (t = 0.19, p<0.05) (Figure 4b). Therefore, we can consider our matching to yield balanced Treated and Control groups of users that allow our ensuing analyses to be controlled on observed covariates.

4.6 Measuring Treatment Effect

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

To quantify the *effect* of treatment, we measure the Relative Treatment Effect (RTE) per outcome measure in every strata, as the ratio of the likelihood of the outcome in the Treated group to that in the Control group [48, 95]. Next, using a weighted average across all the strata, we obtain the mean RTE of the treatment per outcome measure. An outcome RTE greater than 1 would mean that the outcome is greater in the Treated users in the Measurement period — or in our case, that allocating ads according to preferred timing increases the likelihood of ad effectiveness.



(a) Ad Reception **(b) Ad CTR**

Figure 5: Ad Outcomes: (a) Ad Reception and (b) Ad CTR. Treated users show greater ad outcomes on an average.

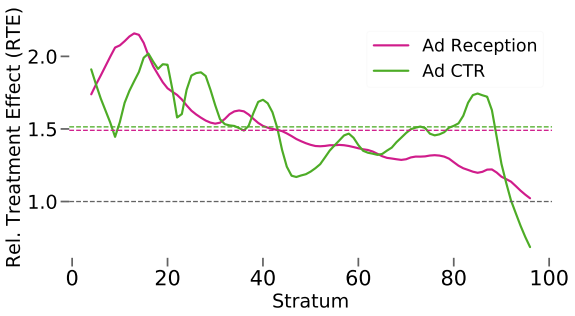


Figure 6: Ad outcomes per matched strata of users.

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

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

4.7 Findings: What is the Effect of Timing Ads?

Figure 5 shows average changes of ad outcomes in Treated and Control users — indicating higher ad outcomes for Treated users. Table 1 reports the differences in outcomes of the Treated and

Control users in terms of RTE, Cohen's d, paired t-test, and KS test. These tests indicate statistically significant outcome differences in Treated and Control users. The high magnitude Cohen's d values (d>0.5) and t-statistic, along with the positive signs indicate that the Treated group shows significantly higher outcomes than the Control group. Additionally, the fact that ad reception shows an RTE of 1.49 and ad CTR shows an RTE of 1.51, indicates that timing ads (treatment) leads to more effective ad outcomes (RTE>1). Figure 6 reveals that the RTE>1 holds for a significant majority of user strata for both ad outcomes.

This reveals that although matched Treated and Control users are very similar on baseline behavior and ad consumption, along with Measurement period's ad exposure and platform activities (as per matching), they show very different receptivity to ads in the Measurement period. Therefore, we draw two inferences. First, the short baseline period of understanding on-platform activities of users can lead us to passively infer *preferred times of ad consumption*. Second, when users are shown ads at preferred times, they are more likely to be receptive to the ads — or they would watch the ads longer, and are more likely to click on them. On the other hand, ads shown during less preferred hours may not only be worse received but may also elicit frustration or seem intrusive to users [27].

5 AIM 2: USER BEHAVIOR AND AD EFFECTIVENESS

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

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

5.1 Comparing Ad Effectiveness by Sessions

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

5.2 Ad Effectiveness by Momentary Behaviors

Ad consumption is known to be a function of people's momentary psychological and behavioral states (Section 2). For instance, ad consumption (and more generally, content consumption) is a function of what an individual does, or how active or tired they are at a particular block of time [13, 102]. As a proxy of behavioral and psychological states, we draw on prior work to operationalize a variety of passively inferred in-the-moment states on the Snapchat platform [96], which are explained below. We first motivate and operationalize each of these in-the-moment behaviors and follow that with our observations with respect to the relationship with ad effectiveness in each. We are particularly looking for insights to recommend ad allocations with respect to preferred times with minimal compromise on ad effectiveness, i.e., 1) when to increase ads during less-preferred times and 2) when to decrease ads during more-preferred times. Table 2 summarizes the differences in high and low ad effective sessions by preferred times of users.

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

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

Interactivity. One way to study user behavior on a social platform is measuring a user's degree of interactiveness in terms of posting, responding, and consuming content [96]. We operationalize interactivity as the ratio of content created to content consumed within

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

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

a session. Here, content creation on the Snapchat platform includes creating stories, posting updates, and sending and replying to chat messages, while content consumption includes viewing others' stories and updates, and browsing through different content within a session. Table 2 shows that for both less-preferred times (t=−3.18) and more-preferred times (t=−5.25), t-statistic is negative. This suggests that interactivity negatively associates with ad effectiveness, or the higher the interactivity of a user in a session, the lower their ad reception. A plausible interpretation of our Interactivity results is that, when users encounter ads on Discover Feed during highly interactive sessions, they might feel particularly disrupted and have specific actions available (e.g. chatting with a friend), motivating them to skip ads. Therefore, our findings recommend to *increase ads* at "low interactivity sessions of less-preferred times", and *decrease ads* at "high interactivity sessions of more-preferred times".

Interaction Diversity. An aspect of social media behavior is the diversity of interactions a user conducts during a session [96]. We operationalize interaction diversity as the standard deviation of time spent in each kind of activity conducted in a session, where in-session activities range across sending or replying to chats, viewing and posting updates, etc [37]. As above, we find statistical significance and negative t-statistics for both less-preferred (t=−5.61) and more-preferred (t=−4.78) times. This suggests that interaction diversity negatively associates with ad effectiveness. Therefore, our findings recommend to *increase ads* at "low interaction diversity sessions of less-preferred times", and *decrease ads* at "high interaction diversity sessions of more-preferred times".

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

Extra-socialness. Platforms such as Snapchat, Instagram, and Facebook also provide features outside conventional forms of social media activities such as posting, chatting, and socially interacting with others, e.g., playing games, using a camera and applying filters or lenses on their photos, etc [37]. We operationalize the ratio of time spent on these activities to the total session duration as extra-socialness of a session. Comparing extra-socialness of different ad reception sessions at less-preferred times, we find no statistical significance, and at more-preferred times, we find t=−1.35 with significance. This indicates that, at more-preferred times, extra-socialness negatively correlates with ad effectiveness, or when a user is interested in non-social platform activities, they are less likely to be receptive to ads, finding them disruptive. Correspondingly, a recommendation would be to decrease ads during high extra-social sessions at more-preferred times of users.

Summary of Insights and Recommendations. The above observations suggest that ad effectiveness positively correlates with duration and negatively correlates with activity, interactivity, and interaction diversity for any time; positively correlates with distractedness and negatively correlates with extra-socialness at more-preferred times. Therefore, recommendations for *increasing ad allocation at less-preferred times* are in sessions with 1) high duration, 2) low activity, 3) low interactivity, and 4) low interaction diversity. On the other hand, recommendations for *decreasing ad allocation at more-preferred times* are in sessions with 1) low duration, 2) high activity, 3) high interactivity, 4) high interaction diversity, 5) low distractedness, and 6) high extra-socialness.

While we study the relationship between on-platform momentary behaviors and ad effectiveness here, approaches to infer momentary behaviors in real-time or apriori are beyond the scope of this study. However, these can be implemented using real-time dynamic rules (e.g., if the current session duration or session interactivity is already higher than the user's median at a given time) or using predictive machine learning techniques [6, 52].

6 AIM 3: EFFICACY OF OUR INSIGHTS

Our Aim 1 and 2 derived insights about ad consumption by ad allocations at preferred hours and on-platform behaviors respectively, we ask how these insights would influence business value? We conduct a simulation experiment of increasing and decreasing ad allocations based on recommendations guided by our first two

research aims. Additionally, we were concerned that intervening in the status quo of the ad allocation process could create concentrated ad allocations, i.e. certain users bearing the burden of high "ad load". Thus, this additional investigation aims to evaluate how simulated interventions might affect the fairness of ad allocations in terms of the concentration of ad load among users. This simulation experiment can inform an ad allocation system about weighing in ad allocations to users who are at extremes of ad exposure to balance the quantity of overall ad exposure across all users on the platform, i.e., a more balanced but effective allocation of ads.

6.1 Simulating Balanced Ad Reallocations

We first quantify the *normalized ad quantity* per user as the ratio of the total number of ads over the total number of pieces of content (or Discover Feed stories) seen by the user. Figure 7a shows the min-max scaled distribution of normalized ad quantity within our Measurement dataset. We identify *high and low ad-exposed users* as the top and bottom quartile of normalized ad quantity. Because these users are at extremes of ad exposure during a particular period, a platform would ideally like to first change the ad quantity of these users to similarly balance out normalized ad quantity across all users. Accordingly, we *simulate* new ad distributions by decreasing the ad quantity of high ad-exposed users and increasing the ad quantity of low ad-exposed users in the following three ways:

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

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

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

6.2 Evaluating Ad Reallocations

We evaluate ad reallocations on the basis of *ad value*, which is a function of how effective ads are in a session. We measure *ad value* as a product of ad reception and the number of ads in a session. Theoretically, ad value would be correlated with actual monetary value generated based on ad effectiveness [72, 105]. We note that our simulations are only within the limits of the observational data, and our measure of ad value assumes a user's ad reception in a session to be the observed value. However, it is likely that the counterfactual ad reception might change if the ads are actually reallocated — which remains unknown unless an actual experiment of ad reallocation is conducted.

Figure 7b shows the distribution of ads across users in multiple simulation strategies. We find that compared to the actual distribution, 1) the baseline simulation decreases the standard deviation by 9%, 2) the simulation by activity decreases by 6.5%, and 3) the simulation by preferred time decreases by 7.9%. Lower standard deviation suggests that all our simulated forms of ad reallocations result in a balanced allocation of ads across users.

Next, Figure 7c shows the distribution of ad value by simulation strategies. We find that compared to overall ad value in the actual distribution: 1) the baseline simulation only marginally increases ad value by 0.07%, whereas 2) the simulation by activity strategy increases ad value by 2.78%, and 3) the simulation by time strategy increases ad value by 7.09%. Both of these percentage changes actually correspond to a significant increase in overall value considering the scale of userbase and volume of data and ads on the platform, e.g., ~250M daily active users on Snapchat [113].

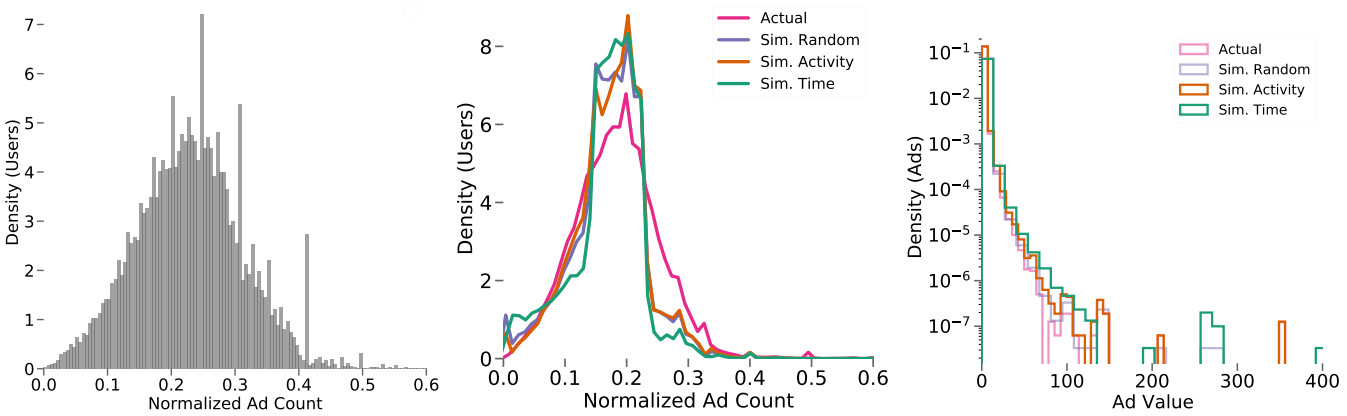
Finally, drawing on permutation test approaches [5], we iterate over the 1,000 permutations of baseline reallocations, to find the probabilities (p-values) of the ad value improvement in the baseline reallocation over the two insight-driven reallocations. We find these probabilities to be zero, suggesting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

7 ROBUSTNESS OF FINDINGS

This section examines the robustness of our findings with respect to the researcher decisions we made in our study. First, we conduct methodological robustness tests on parametric choice and approach in our study design. Then, we theoretically contextualize our definition of "ad effectiveness" measures with respect to how it is defined traditionally in the literature — a success would provide criterion and construct validity to our study. Together, a convergence in findings along with theoretical grounding would ensure robustness and validity of our findings [60].

7.1 Methodological Robustness

Binarizing Treatment Dosage Thresholds Recall that our study design relies on chosen threshold of binarizing treatment and no-treatment (Section 4). Therefore, we test if our findings hold robust for any other thresholds of treatment. For this, we vary the threshold of treatment dosage and re-conduct the entire analyses on measuring treatment effects (Aim 1), including matching and computing differences in outcomes for matched Treated and Control users. We vary the threshold parameter α in such a way that



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

Figure 7: Distribution of a) normalized ad count by user in the actual dataset, b) normalized ad count as per simulations of ad reallocations: the density plot for each simulate reallocation is thinner in width compared to the actual distribution suggesting a more balanced ad allocation across users, c) ad value as per simulations: overall ad value is highest for simulated by time reallocation (Sim. Time) followed by simulated by behavioral activty based reallocation (Sim. Activity).

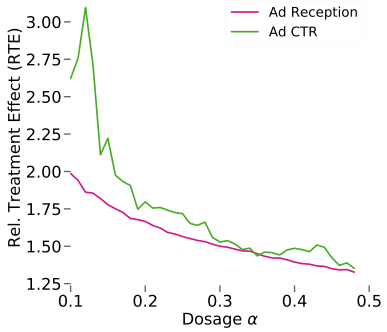


Figure 8: RTE with varying dosage α. For each α, treatment is top $\alpha*100$ percentile of dosage and no-treatment is bottom $\alpha * 100$ percentile of dosage.

dosage (cosine similarity between Baseline ad reception and Measurement ad allocatino) in the top $\alpha * 100$ percentile is considered treatment, and that in the bottom $\alpha * 100$ percentile is considered no-treatment. Figure 8 plots the change in RTE of ad effectiveness measures, with respect to changing the threshold dosage α. We find that the RTE of both ad effectiveness measures remain greater than 1 (along with statistical significance as per t-test and effective size), indicating that ads were more effective on Treated users or users who were shown ads at their preferred hours. We also find a roughly monotonic decrease in RTE with respect to increasing α, suggesting the greater the similarity of ad allocations with people's preferred hours, the greater is the likelihood of ad effectiveness.

Using Treatment Dosage as a Continuous Variable. Another component of our work includes the decision to estimate the outcomes by binarizing the treatment dosage. Such an approach not only better serves interpretability purposes but also emulates conventional RCT or experimental approaches where one group is treated (e.g., drug) and the other group is not (e.g., placebo). Though less likely, binarizing treatment might however introduce new biases in the analyses leading to misleading findings (e.g. a drug in low dosage may not be as effective as it is in high dosage [40]). Therefore, we test the findings if the treatment is considered as a continuous variable.

For this, we build a linear regression model that uses all the 42 covariates in our dataset, along with the treatment as independent variables and the ad reception as the dependent variable. While this approach is not particularly "causal", it allows us to infer the relationship of the treatment and the outcomes. We eliminate correlated features using variance inflation factor (VIF) (threshold=10) [19, 70]. The regression model shows an adjusted R^2 of 0.87, and Table 3 reports the standardized coefficients of relevant variables. In particular, we find that the treatment dosage shows the greatest magnitude with a positive coefficient of 0.22 (p<0.05). Likewise, repeating the same experiment with ad CTR as dependent variable leads to similar signs of coefficients. Together, the findings from our regression analysis aligns with our matched and binarized treatment analysis that *treatment* (or showing ads during preferred hours) leads to a greater likelihood of ad effectiveness.

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

7.2 Contextualizing within the Literature

As a final robustness check, we compare our results with that found in previous literature. Traditionally, ad effectiveness is defined as whether an individual buys a commodity following exposure to an ad [13]. If our observations of ad effectiveness match prior literature, we view that as criterion validity to our measures of ad effectiveness

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

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

and construct validity to our study [75]. We test ad effectiveness in two ways: in terms of the time of day and as day of the week.

7.2.1 Time of the day. We construct our first hypothesis based on prior work comparing ad effectiveness at days and nights [102] that: *ad effectiveness is higher during the daytime compared to evenings or nights*. In our work, we first bucket a user's local time into day (6 AM–6 PM) and night (6 PM–6 AM). While we also attempted to build more granular buckets or even look at ad effectiveness over more continuous forms of time, we find effects are generally washed out given the across-user variability in ad reception, also evident in our preliminary analysis referring to Figure 2c, particularly in the bottom-most/"All" user row. Instead, binary buckets (day and night) provide us the opportunity to compare and contrast users' receptivity to ads between broader timespans of a day.

For both Treated and Control groups, we conduct paired t-tests between a user's ad outcomes during the day and during the night. Table 4 reports the differences in ad reception by time of the day. First, understandably, ad effectiveness for Treated users is higher than Control users as also reflected in previous analyses (Section 4). Next, for both Treated and Control users, ad effectiveness is higher during the day than night with statistical significance and large t-statistics, supporting prior research on ad effectiveness [102].

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

We statistically compare ad effectiveness on weekends and weekdays, as reported in Table 4. First, ad effectiveness is higher for Treated than Control users. Next, within Control users, we find that both forms of ad outcome are higher during the weekends than on weekdays. In contrast, in the case of Treated users, ad reception during the weekdays and weekends do not differ with statistical significance. This indicates support for our hypothesis in the case of Control users, and lack of support in the case of Treated users. These findings suggest that users in the Treated group already see ads at their preferred time, and as a result the weekend/weekday effect is diminished. Therefore, whether timing ads likely plays a stronger role than day of the week effect — future research may be able to shine more light on this.

8 DISCUSSION

8.1 Theoretical Implications

Our work opens up discussions on understanding ad allocation and consumption in new forms of media. Traditional media (television, newspapers) drive ads dedicated to audience groups, and ad effectiveness typically measures the amount of product purchases (by ad influence). In recent times, not only with the increasing use of online social platforms, but also with the ubiquity of smart and personal devices, ads are allocated in various novel ways. The affordances not only enable platforms to customize and allocate ads in a personalized fashion, but also provide users with choices to skip and ignore ads. Importantly, negative perception towards ads can cause user fatigue and exacerbate their perception of the platform [12]. This calls for a need to better understand ad consumption and accordingly design robust and dynamic ad allocation strategies. Our work reveals that when ads are allocated in a better fashion by accounting for user- and context-centric factors (e.g., time and platform engagement), users could be more receptive to ads, which corresponds to prior observations regarding intrusiveness and likeability of ads [50, 63].

We have couched our observations in theories from marketing science, psychology, and cognitive science. Our work augments prior research which largely studied how the content of ads matters in changing user experience on online platforms [4, 17]. Our work provides valuable insights regarding the importance of context and momentary factors in understanding ad effectiveness. In particular, our observations suggest that timing ads is a factor that cannot be ignored when allocating ads on social media.

Along similar lines, our work also reveals how blanket approaches of ad allocation or approaches based on average user behavior may not be as effective. For instance, these approaches typically assume a linear relationship between a user's time spent in a session and the number of ads shown: ads are shown after a fixed duration or after showing a fixed number of pieces of content. However, these approaches ignore the cognitive state of users which can vary due to time of day effects or due to users' daily routines, or even momentary psychological states such as feeling social or excited at a particular moment [33, 102]. Instead, our work finds that when ads are allocated by accounting for these factors, ad effectiveness is higher without compromising the user activities on the platform.

8.2 Practical and Design Implications

Individual-centric Implications. Our work has implications for making responsible advertising — advertising that aims to not only increase platform revenue, but also minimizes user dissatisfaction caused by ads, helping to keep the users better engaged with the platform [103]. Our work contributes towards the niche aspect of "preferred timings" and provides an approach to balance ad effectiveness and user experience. Our approach can help to minimize the privacy intrusions generally associated with targeted advertising: We can reduce the use of profiling to target ads, and obtain preferred ad timing on de-identified features and short-term data.

Because users typically dislike ads and do not like to share their data with platforms for ad targeting [73], they often use tools such as ad-blocking and private browsing that do not share cookies and browser history with advertisers. Some platforms disallow these

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

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

privacy-preserving practices, forcing the user to trade off their data in order to access the content. Potentially, users may be more comfortable sharing only their momentary (session-level) data, and our work shows that platforms can make effective ad allocations by only using these minimal, momentary user data.

Further, the efficacy of recommending content on the basis of time and context we have shown, suggests design implications that take user agency and user preferences into account [103]. Recently, the HCI community has demonstrated the value of user-contributed preferences of notifications and interruptions [66, 79, 103]. In line with this, platforms could ask users which times of the day they are more inclined to view ads, and allocate ads accordingly. This approach will require more insights in potential ways users could game the system (users might provide preferred times of ads when they would likely not visit the platform).

Platform-centric Implications. Our methodology allows platforms to allocate ads in an effective, fair, and less-intrusive way. A recent survey revealed three major categories of ad dissatisfaction are intrusiveness, annoyance, and disruptiveness of ads [108]. Users often use ad blockers and other tools that prevent ads on online platforms [30, 71] — these approaches raise nuanced questions surrounding the sustainability of platforms surviving on ad-driven business models [4]. Consequently, to protect user base and minimize ad-based interruptions, some platforms are moving away from ad-based models to some form of subscription-based models [31, 109]. However, such models have their own caveats, such as inequity of information access on the internet, and online services could become a function of an individual's ability to pay. Our work suggests somewhat of a middle-ground: by optimizing ad timings and allocations when users are less likely to feel interrupted, platforms can consistently provide equitable content access and experience to users, and better sustain the ad revenue ecosystem, with less user dissatisfaction.

Towards Fewer Ads. Our work has implications towards optimizing other forms of ad allocations on social media, including ad spacing and ad loading. One can draw an insight that if we can allocate ads optimally in an effective fashion, we can plausibly reduce the overall quantity of ads if certain revenue goals are already achieved with smaller quantity of, but better-allocated (timed) ads. This can help minimize practices such as non-skippable ads (ads which cannot be skipped) or forced ads (ads which prevent any content consumption without being watched). Solutions of minimal ad allocations would be well-appreciated by the users, improving the general user experience, potentially leading to higher user retention on the platform [65]. A better ad allocation strategy provides a method for platforms to judiciously serve effective ads. As a result, such platforms can become more attractive to users.

Small Data and Privacy-sensitive Approaches. Research highlights several biases in ad delivery [3, 56, 85]. For instance, demography and inferred-user interest-based targeting can be deemed privacy-intrusive, unethical, and surveillance-promoting [41, 82, 88]. In contrast, our work shows a novel means to increase ad effectiveness that does not lean on these critiqued approaches. Our approach only requires short-term user data (e.g. a few weeks) instead of using long-term historical data. Long-term data do not only increase privacy concerns, but are also less robust to changes in both human behavior and platform affordances [9]. So, small-data-driven approaches that do not compromise on the user experience can open up new opportunities in ad and content recommendation.

Other Content Recommendations. While we primarily focus on ads, our work also has implications for other types of content. Our work provides general insights for when to show content to users. This could inform design strategies for recommendations and notifications for preferred user content. Prior work showed the value of context-aware recommendations for improving user engagement on mobile platforms [47, 66]. The on-platform behaviors we studied (particularly in Section 5) can guide designing such recommendations that take a user's momentary state into account.

Implications for Experimental Approaches. As we noted earlier, causal effects are best studied with experimental approaches, which however, come with risks, e.g., certain treatments (design changes) may affect the perception of users and impact user retention. Moreover, in the case of continuous treatments, it is often difficult to determine the appropriate dosage to experiment on. Our study adopts a quasi-experimental design to show that a particular treatment (timing ads) can be effective. Our computational framework also quantifies "preferred timings" based on observed ad outcomes in a small time period. Therefore, our findings can help to formulate appropriate parameters (or dosage cutoffs as shown in Section 7) to conduct careful experimental studies to verify and adopt design changes on platforms [42, 71].

Substitute or a Complement to the Existing Ad Allocation System? Lastly, we raise a critical point. We conduct our study on a system already optimized (in some form) for ad allocations. Therefore, our effects may be even bigger if we had a different baseline. Arguably, our study only builds on the top of the existing

system which might already be privacy-intrusive and be using inferred user interests based targeting — the very points on which we discuss several implications above. In this regard, it indeed remains unexplored whether momentary and context-driven features will be adopted in practice to improve ad effectiveness. There is a risk that companies will stack profile and momentary state approaches (instead of replacing the profile-based approaches), which could potentially be more privacy invasive. However, our study encourages considering and evaluating these alternative strategies that conduct responsible and non-invasive ad allocations. It would be immensely insightful if future research suggests that we can (or cannot) significantly minimize or even eliminate any sort of profiling and demographic based targeting approaches. Ad targeting is coming under increased scrutiny, and as companies and governments are putting in place more privacy restrictions [107], our approach can help future-proof the ecosystem of ad-based platforms. Overall, an important takeaway of this work illustrates the importance and feasibility of building complementary methodologies which simultaneously consider a user's privacy and optimize for user experience and business value for long-term sustainability.

8.3 Ethical and Privacy Implications

We note that our work bears ethical implications. Our work is predominantly motivated by the idea that we might move away from traditional forms of user profiling and using static demographic and trait-based information that content recommendation algorithms infer, which can be biased and unfair [41, 82]. However, this work can be (mis)used to conduct new forms of user profiling on people's online behavior. Online platforms could (mis)use our approach to conduct newer and plausibly unknown forms of biased and intrusive ad targeting, e.g., if these algorithms incorporate not only "who someone is", but also "what someone does when". Pandit and Lewis described, "the use of personal data is a double-edged sword that on one side provides benefits through personalisation and user profiling, while the other raises several ethical and moral implications that impede technological progress" [77]. Therefore, we need to consider balancing the costs and benefits of these approaches to implement them in ethical and privacy-preserving fashion. While arguably anonymized and on-platform in-the-moment behavior is more ethical and less biased compared to demographic, static, and prior-assumptions based stratification, we also recognize the possibility of expectation mismatches between users' self-conceptualization of their data and inferences on their data without awareness [28]. For example if personalized ad allocations start working even better (by using momentary and contextual data), ads may seem "creepier" [56, 74] — as Malheiros et al. noted, "too personalized" ads can although catch more attention, but could also elicit discomfort about the personalization [61].

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

8.4 Limitations and Future Directions

Our work has limitations, some of which also suggest interesting future directions. We do not take content (e.g., what an ad is about) into account. While our work is a step towards understanding the role of context and momentary features, we note that future research can incorporate content to examine the interplay between context and content-centric factors in explaining online ad effectiveness. Additionally, our study functions within the limits of the existing ad allocation system on the platform. Because all users in the Treated and Control groups were allocated ads by the same system, the effects of ad allocation are likely balanced or washed out in the large scale of data we worked with. Therefore, despite the inaccuracies and limitations of ad recommender system, we consider our findings to remain valid, and our claim is reinforced by the empirical robustness and theoretical validity tests conducted in Section 7. We also note that our data included a diversity of ads spanning a wide range of costs and types across the Treated and Control datasets. As a proxy for major group differences in ads, a t-test on ad distributors of Tr and Ct groups reveals no significant difference (p>0.1). Therefore, although we assumed platform-centric factors to apply similarly to all users and discrepancies to balance across groups through the scale of the data, this assumption may not fully hold. Future work can include these confounds and control for the "goodness" of recommender algorithms if such metrics are available.

Our quasi-experimental approach only accounts for observed factors on the platform. Like any observational study, we cannot infer *true causality*. Future experimental studies based on our methodology and insights from our study can help confirm the validity and applicability of our findings. Similarly, we only quantify *observed* ad effectiveness, and cannot estimate the efficacy of the ads in terms of whether the users actually bought or used the products in the ads [67]. Future studies can enroll small representative samples of the population and conduct experimental studies that also incorporate the offline element of effectiveness of online ads. Future work can also study the "why"-s related to whether users like a particular kind of ad allocation versus the other [58]. Along the same lines, future studies can examine if providing explanation to ad allocations make users more (or less) conducive to ads [49]. We also cannot claim the generalizability of our findings on other platforms and other forms of ad allocations, which can be explored in future work. Our work builds the foundation for incorporating context and momentary behaviors in ad allocations, which can be extended in the future to other forms of content recommendation systems and problem settings.

9 CONCLUSION

This paper examined ad consumption on online social platforms, particularly on the Snapchat Discover Feed. We conducted a quasi-experimental study on three months of longitudinal data of 100K Snapchat users. We split the longitudinal timeline of each user into

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

ACKNOWLEDGMENTS

This work was conducted while Saha, Vincent, and Chowdhury were at Snap Research. We thank Andrés Monroy-Hernandez, Dong Whi Yoo, and Vedant Das Swain for their valuable feedback.

REFERENCES

[1] 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of clinical epidemiology* 67, 3 (2014), 267–277.
[2] Paige Alfonzo. 2019. *Mastering mobile through Social Media: Creating engaging content on Instagram and Snapchat.* ALA TechSource.
[3] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
[4] Maximilian Altmeyer, Kathrin Dernbecher, Vladislav Hnatovskiy, Marc Schubhan, Pascal Lessel, and Antonio Krüger. 2019. Gamified Ads: Bridging the Gap Between User Enjoyment and the Effectiveness of Online Ads. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
[5] Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. 2008. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 7–15.
[6] Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, and Anind K Dey. 2016. Modeling and understanding human routine behavior. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 248–260.
[7] Rajeev Batra and Douglas M Stayman. 1990. The role of mood in advertising effectiveness. *Journal of Consumer research* 17, 2 (1990), 203–214.
[8] Russell W Belk. 1974. An exploratory assessment of situational effects in buyer behavior. *Journal of marketing research* 11, 2 (1974), 156–163.
[9] Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.
[10] Giorgio Brajnik and Silvia Gabrielli. 2010. A review of online advertising effects on the user experience. *International Journal of Human-Computer Interaction* 26, 10 (2010), 971–997.
[11] Jack W Brehm. 1966. A theory of psychological reactance. (1966).
[12] Laura Frances Bright and Kelty Logan. 2018. Is my fear of missing out (FOMO) causing fatigue? Advertising, social media fatigue, and the implications for consumers and brands. *Internet Research* (2018).
[13] Fred E Bronner, Jasper R Bronner, and John Faasse. 2007. In the mood for advertising. *International Journal of Advertising* 26, 3 (2007), 333–355.
[14] Bobby J Calder, Edward C Malthouse, and Ute Schaedel. 2009. An experimental study of the relationship between online engagement and advertising effectiveness. *Journal of interactive marketing* 23, 4 (2009), 321–331.
[15] Jean-Louis Chandon, Mohamed Saber Chtourou, and David R Fortin. 2003. Effects of configuration and exposure levels in responses to web advertisements. *Journal of Advertising Research* 43, 2 (2003), 217–229.
[16] Jean-Charles Chebat, Francois Limoges, and Claire Gelinas-Chebat. 1997. Effects of circadian orientation, time of day, and arousal on consumers' depth of information processing of advertising. *Perceptual and motor skills* (1997).
[17] Henriette Cramer. 2015. Effects of ad quality & content-relevance on perceived content quality. In *proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2231–2234.
[18] Aron Culotta. 2014. Estimating county health statistics with Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1335–1344.
[19] Vedant Das Swain et al. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *Proc. ACM IMWUT* 3, 4, Article 130 (2019), 27 pages. https://doi.org/10.1145/3369828
[20] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2098–2110.
[21] Marco de Sa, Vidhya Navalpakkam, and Elizabeth F Churchill. 2013. Mobile advertising: evaluating the effects of animation, user and content relevance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2487–2496.
[22] Peter Doyle and John Saunders. 1990. Multiproduct advertising budgeting. *Marketing Science* 9, 2 (1990), 97–113.
[23] Robert H Ducoffe. 1995. How consumers assess the value of advertising. *Journal of Current Issues & Research in Advertising* 17, 1 (1995), 1–18.
[24] Robert H Ducoffe. 1996. Advertising value and advertising on the web-Blog@ management. *Journal of advertising research* 36, 5 (1996), 21–32.
[25] Robert H Ducoffe and Eleonora Curlo. 2000. Advertising value and advertising processing. *Journal of Marketing Communications* (2000).
[26] Julie A Edell and Marian C Burke. 1984. The moderating effect of attitude toward an ad on ad effectiveness under different processing conditions. *ACR North American Advances* (1984).
[27] Steven M Edwards, Hairong Li, and Joo-Hyun Lee. 2002. Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads. *Journal of advertising* 31, 3 (2002), 83–95.
[28] Casey Fiesler, Cliff Lampe, and Amy S Bruckman. 2016. Reality and perception of copyright terms of service for online content creation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 1450–1461.
[29] Meryl Paula Gardner. 1985. Mood states and consumer behavior: A critical review. *Journal of Consumer research* 12, 3 (1985), 281–300.
[30] Kiran Garimella, Orestis Kostakis, and Michael Mathioudakis. 2017. Ad-blocking: A study on performance, privacy and counter-measures. In *Proceedings of the 2017 ACM on Web Science Conference*. 259–262.
[31] Bob Gilbreath. 2017. Rise of Subscriptions and the Fall of Advertising: *https://medium.com/the-graph/rise-of-subscriptions-and-the-fall-of-advertising-d5e4d8800a49*.
[32] Scott A Golder and Michael W Macy. 2011. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333, 6051 (2011), 1878–1881.
[33] Kendall Goodrich. 2013. Effects of age and time of day on Internet advertising outcomes. *Journal of Marketing Communications* (2013).
[34] Samuel D Gosling, Adam A Augustine, Simine Vazire, Nicholas Holtzman, and Sam Gaddis. 2011. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking* 14, 9 (2011), 483–488.
[35] Kelley Gullo, Jonah Berger, Jordan Etkin, and Bryan Bollinger. 2019. Does time of day affect variety-seeking? *Journal of Consumer Research* 46, 1 (2019), 20–35.
[36] Qi Guo, Eugene Agichtein, Charles LA Clarke, and Azin Ashkan. 2009. In the mood to click? Towards inferring receptiveness to search advertising. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 319–324.
[37] Hana Habib, Neil Shah, and Rajan Vaish. 2019. Impact of Contextual Factors on Snapchat Public Sharing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[38] Edward L Hannan. 2008. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC* (2008).
[39] Curt Haugtvedt, Richard E Petty, John T Cacioppo, and Theresa Steidley. 1988. Personality and ad effectiveness: Exploring the utility of need for cognition. *ACR North American Advances* (1988).
[40] Miguel A Hernan and James M Robins. 2010. *Causal inference.*
[41] Ben Hutchinson and Margaret Mitchell. 2019. 50 years of test (un) fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 49–58.
[42] Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge.
[43] Garrett A Johnson, Randall A Lewis, and Elmar I Nubbemeyer. 2017. Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of*

Marketing Research 54, 6 (2017), 867–884.
[44] Jukka Jouhki, Epp Lauk, Maija Penttinen, Niina Sormanen, and Turo Uskali. 2016. Facebook's emotional contagion experiment as a challenge to research ethics. *Media and Communication* 4 (2016).
[45] Parisa Kaghazgaran, Maarten Bos, Leonardo Neves, and Neil Shah. 2020. Social Factors in Closed-Network Content Consumption. *CIKM* (2020).
[46] Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A Konstan, and Paul Schrater. 2015. "I like to explore sometimes" Adapting to Dynamic User Novelty Preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 19–26.
[47] Komal Kapoor, Karthik Subbian, Jaideep Srivastava, and Paul Schrater. 2015. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. 233–242.
[48] 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.
[49] Tami Kim, Kate Barasz, and Leslie K John. 2019. Why am I seeing this ad? The effect of ad transparency on ad effectiveness. *Journal of Consumer Research* 45, 5 (2019), 906–932.
[50] Yoojung Kim, Mihyun Kang, Sejung Marina Choi, and Yongjun Sung. 2016. To click or not to click? Investigating antecedents of advertisement clicking on Facebook. *Social Behavior and Personality: an international journal* 44, 4 (2016), 657–667.
[51] Gary King, Richard Nielsen, et al. 2016. Why propensity scores should not be used for matching. (2016).
[52] Farshad Kooti, Karthik Subbian, Winter Mason, Lada Adamic, and Kristina Lerman. 2017. Understanding short-term changes in online activity sessions. In *Proceedings of the 26th International Conference on World Wide Web Companion*.
[53] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. (2013).
[54] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24 (2014), 8788–8790.
[55] Hemank Lamba and Neil Shah. 2019. Modeling dwell time engagement on visual multimedia. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1104–1113.
[56] Zhou Li, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang. 2012. Knowing your enemy: understanding and detecting malicious web advertising. In *Proceedings of the 2012 ACM conference on Computer and communications security*. 674–686.
[57] Q Vera Liao, Wai-Tat Fu, and Sri Shilpa Mamidi. 2015. It is all about perspective: An exploration of mitigating selective exposure with aspect indicators. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*. 1439–1448.
[58] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2119–2128.
[59] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing behavioral stages from social media data. In *CSCW*.
[60] Xun Lu and Halbert White. 2014. Robustness checks and robustness tests in applied economics. *Journal of econometrics* 178 (2014), 194–206.
[61] Miguel Malheiros, Charlene Jennett, Snehalee Patel, Sacha Brostoff, and Martina Angela Sasse. 2012. Too close for comfort: A study of the effectiveness and acceptability of rich-media personalized advertising. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 579–588.
[62] En Mao and Jing Zhang. 2015. What drives consumers to click on social media ads? The roles of content, media, and individual factors. In *2015 48th Hawaii International Conference on System Sciences*. IEEE, 3405–3413.
[63] En Mao and Jing Zhang. 2017. What affects users to click on display ads on social media? The roles of message values, involvement, and security. *Journal of Information Privacy and Security* 13, 2 (2017), 84–96.
[64] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3025–3034.
[65] Jack Marshall. 2016. How to Persuade Consumers to Disable Ad Blockers: *https://www.wsj.com/articles/how-to-persuade-consumers-to-disable-ad-blockers-1469541611*.
[66] Akhil Mathur, Nicholas D Lane, and Fahim Kawsar. 2016. Engagement-aware computing: Modelling user engagement from mobile contexts. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 622–633.
[67] Daniel McDuff, Rana El Kaliouby, Jeffrey F Cohn, and Rosalind W Picard. 2014. Predicting ad liking and purchase intent: Large-scale analysis of facial responses to ads. *IEEE Transactions on Affective Computing* 6, 3 (2014), 223–235.
[68] Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017).
[69] Jacob Metcalf and Kate Crawford. 2016. Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society* 3, 1 (2016), 2053951716650211.
[70] Jeremy Miles. 2014. Tolerance and variance inflation factor. *Wiley StatsRef: Statistics Reference Online* (2014).
[71] Ben Miroglio, David Zeber, Jofish Kaye, and Rebecca Weiss. 2018. The effect of ad blocking on user engagement with the web. In *Proceedings of the 2018 World Wide Web Conference*. 813–821.
[72] Bruce John Morrison and Marvin J Dainoff. 1972. Advertisement complexity and looking time. *Journal of marketing research* 9, 4 (1972), 396–400.
[73] Ngoc Thi Nguyen, Agustin Zuniga, Hyowon Lee, Pan Hui, Huber Flores, and Petteri Nurmi. 2020. (M) ad to See Me? Intelligent Advertisement Placement: Balancing User Annoyance and Advertising Effectiveness. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2020).
[74] Katie O'Donnell and Henriette Cramer. 2015. People's perceptions of personalized ads. In *Proceedings of the 24th International Conference on World Wide Web*. 1293–1298.
[75] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.
[76] Alexandra Olteanu, Onur Varol, and Emre Kiciman. 2017. Distilling the outcomes of personal experiences: A propensity-scored analysis of social media. In *Proc. CSCW*.
[77] 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*.
[78] Pew. 2019. pewinternet.org/fact-sheet/social-media.
[79] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017).
[80] Lin Qiu, Han Lin, Jonathan Ramsay, and Fang Yang. 2012. You are what you tweet: Personality expression and perception on Twitter. *Journal of research in personality* 46, 6 (2012), 710–718.
[81] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. [n.d.]. Our twitter profiles, our selves: Predicting personality with twitter.
[82] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 469–481.
[83] Vennila Ramalingam, B Palaniappan, N Panchanatham, and S Palanivel. 2006. Measuring advertisement effectiveness—a neural network approach. *Expert systems with applications* 31, 1 (2006), 159–163.
[84] Pei-Luen Patrick Rau, Qingzi Liao, and Cuiling Chen. 2013. Factors influencing mobile advertising avoidance. *International Journal of Mobile Communications* 11, 2 (2013), 123–139.
[85] Filipe N Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P Gummadi, and Elissa M Redmiles. 2019. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 140–149.
[86] Steven Richmond. 2018. How Snapchat makes money. *Investopedia. Elérés* (2018).
[87] Helen Robinson, Anna Wysocka, and Chris Hand. 2007. Internet advertising effectiveness: the effect of design on click-through rates for banner ads. *International Journal of Advertising* 26, 4 (2007), 527–541.
[88] Christian Rohrer and John Boyd. 2004. The rise of intrusive online advertising and the response of user experience research at Yahoo!. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. 1085–1086.
[89] Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* 100, 469 (2005), 322–331.
[90] Adam Sadilek, Henry A Kautz, and Vincent Silenzio. 2012. Modeling Spread of Disease from Social Interactions.. In *International Conference on Weblogs and Social Media (ICWSM)*.
[91] Koustuv Saha et al. 2019. Imputing Missing Social Media Data Stream in Multisensor Studies of Human Behavior. In *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII 2019)*.
[92] Koustuv Saha, Larry Chan, Kaya De Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 95.
[93] Koustuv Saha, Ted Grover, Stephen Mattingly, Vedant Das Swain, Pranshu Gupta, Gonzalo J Martinez, Pablo Robles-Granda, Gloria Mark, Aaron Striegel, and Munmun De Choudhury. 2021. Person-Centered Predictions of Psychological Constructs with Social Media Contextualized by Multimodal Sensing. *PACM IMWUT* (2021).

[94] Koustuv Saha and Amit Sharma. 2020. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. In *ICWSM*.

[95] 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*.

[96] 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*.

[97] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *ICDCS*.

[98] Ann E Schlosser, Sharon Shavitt, and Alaina Kanfer. 1999. Survey of Internet users' attitudes toward Internet advertising. *Journal of interactive marketing* 13, 3 (1999), 34–54.

[99] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791.

[100] Yi Shen, Heshan Sun, Cheng Suang Heng, and Hock Chuan Chan. 2020. Facilitating Complex Product Choices on E-commerce Sites: An Unconscious Thought and Circadian Preference Perspective. *Decision Support Systems* (2020), 113365.

[101] Elizabeth A Stuart. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, 1 (2010), 1.

[102] Gerard J Tellis, Rajesh K Chandy, Deborah MacInnis, and Pattana Thaivanich. 2005. Modeling the microeffects of television advertising: Which ad works, when, where, for how long, and why? *Marketing Science* (2005), 359–366.

[103] Catherine E Tucker. 2014. Social networks, personalized advertising, and privacy controls. *Journal of marketing research* 51, 5 (2014), 546–562.

[104] Daniel Tunkelang. 2018. Are Ads Really That Bad?: *https://medium.com/@dtunkelang/are-ads-really-that-bad-1c3d315f6689*.

[105] Richard Vaughn. 1980. How advertising works: A planning model. *Journal of advertising research* (1980).

[106] Aku Visuri, Simo Hosio, and Denzil Ferreira. 2017. Exploring mobile ad formats to increase brand recollection and enhance user experience. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. 311–319.

[107] Zack Whittaker. 2020. Apple's iOS 14 will give users the option to decline app ad tracking: *https://techcrunch.com/2020/06/22/apple-ios-14-ad-tracking*.

[108] Max Willer. 2018. New Data on Why People Hate Ads: Too Many, Too Intrusive, Too Creepy: *https://www.vieodesign.com/blog/new-data-why-people-hate-ads*.

[109] Max Willer. 2019. The Advertising Industry Has a Problem: People Hate Ads: *https://www.nytimes.com/2019/10/28/business/media/advertising-industry-research.html*.

[110] Stephan Winter, Ewa H Maslowska, and Anne L Vos. 2020. The effects of trait-based personalization in social media advertising. *Computers in Human Behavior* (2020), 106525.

[111] Lori D Wolin, Pradeep Korgaonkar, and Daulatram Lund. 2002. Beliefs, attitudes and behaviour towards Web advertising. *International Journal of Advertising* 21, 1 (2002), 87–113.

[112] Seounmi Youn and Seunghyun Kim. 2019. Understanding ad avoidance on Facebook: Antecedents and outcomes of psychological reactance. *Computers in Human Behavior* 98 (2019), 232–244.

[113] Zephoria. 2019. *https://zephoria.com/top-10-valuable-snapchat-statistics/*.

[114] Justine Zhang, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil. 2020. Quantifying the Causal Effects of Conversational Tendencies. *PACM HCI (CSCW)* (2020).

APPENDIX

Table A1: List of the covariates used in our study.

Baseline Ad Reception	Avg. Direct Snap in Chat Send Count	Avg. Direct Snap Reply Send Count
Total Snap Time Seconds (Baseline)	Avg. Filter Lens Swipes Count	Avg. Discover Feed Ads Count
Total Snap Time Seconds (Measurement)	Avg. Filters Swipes Count	Avg. Full Regular Ad Views Count
Avg. Number of App Opens	Avg. Snap Save Count	Avg. Hour of Day
Avg. Number of App Opens from Notifications	Avg. Snap Screenshot Count	Avg. Day of Week
Avg. Session Time	Avg. Group Snap Send Count	Avg. Session Duration
Avg. Chat View Count	Avg. Story Post Count	Avg. Activity
Avg. Chat Send Count	Avg. Friend Feed Friend Stories View Count	Avg. Content Consumption
Avg. Chat Screenshot Count	Avg. Story Delete Count	Avg. Content Curation
Avg. Direct Snap Create Count	Avg. Story Save Count	Avg. Interactivity
Avg. Direct Snap Send Count	Avg. Fully Watched Ads Count	Avg. Interaction Diversity
Avg. Direct Snap View Count	Avg. Discover Feed View Count	Avg. Distractedness
Avg. Direct Snaps from Chat Feed Send Count	Avg. Full Ad Views in Discover Feed Count	Avg. Extra-Socialness
Avg. Direct Snap with Camera Send Count	Avg. Non-full Ad Views in Discover Feed Count	

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 κ 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 χ^2-square test, which reveals $\chi^2 = \texttt{NaN}$ and $\texttt{p} = \texttt{NA}$, 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 $\textbf{Model}_1$.
- *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 **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

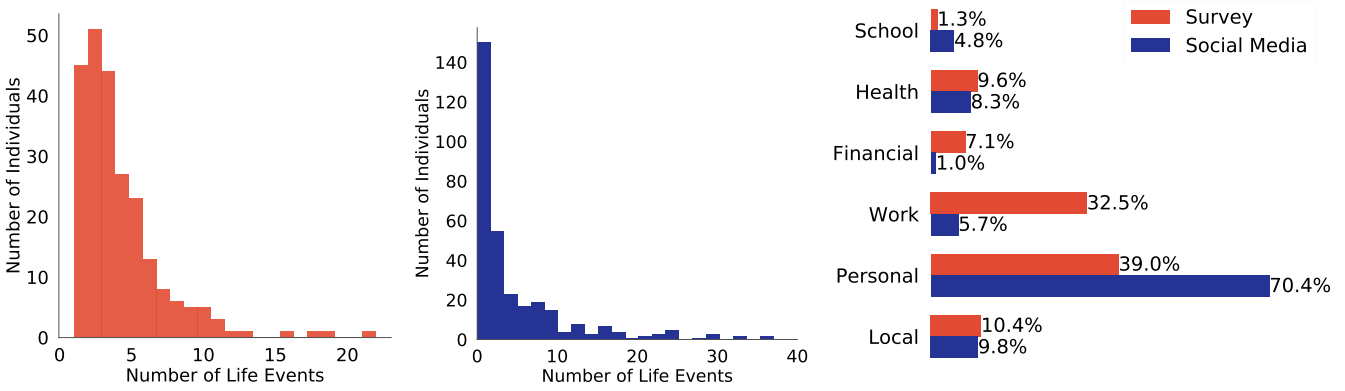
We are now interested to understand how individuals describe life event occurrences on their Facebook timelines. We investigate relationships between social media posts that were temporally similar to life events self-reported on surveys. In particular, for each individual, we look for events that were 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. $\texttt{Model}_1$ examines the factors associated with life events reportage on *both of* against on *one of* the modalities (ref: Table 6). $\texttt{Model}_1$ 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 ($\texttt{e}=-0.33$) indicating that significant events are less likely to be reported on both modalities. Anticipated events are likely to be reported on both ($\texttt{e}=0.16$); 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 $\texttt{R}^2=0.77$, $\chi^2(34)=1785.83$ with $\texttt{p}<0.001$, 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 ($\texttt{e}=-0.38$) 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 ($\texttt{e}=0.73$) — indicating that individuals with greater agreeableness have a likelihood to self-disclose life events on social media. Similarly, extraversion shows a positive coefficient ($\texttt{e}=0.13$). 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 ($\texttt{e}=-0.05$), 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 ($\texttt{e}=0.45$) and anticipation ($\texttt{e}=0.45$) 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, χ^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, χ^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 DMello, 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 (**Model$_1$**) and divergence models (**Model$_2$**) as studied in Section 5.3, we also run a third kind of logistic regression models on the entire data of D_T, such that:

- **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.
- **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, $\texttt{Model}_2$ 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 **Model$_{3a}$** and **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 **Model$_{3a}$** and **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.

	$Model_{3a}$		$Model_{3b}$			$Model_{3a}$		$Model_{3b}$	
Demographic/Trait	**Coeff.**	**p**	**Coeff.**	**p**	**Event Attribute**	**Coeff.**	**p**	**Coeff.**	**p**
Age	0.02	*	−0.04	***	Valence: Positive	0.39	***	−0.19	
Gender: Male	−0.89	*	0.43	***	Significance	−1.26	***	0.87	***
Born in US: Yes	−0.35		0.22		Recency	−1.88		1.57	***
Education: H. School	−0.02		1.16	***	Anticipated	0.19	*	−0.16	
Education: College	−0.04		1.19	*	Intimacy	−0.78	***	0.36	***
Education: Grad School	0.22		0.94	**	Scope	−0.84	***	0.44	**
Education: Doctoral	0.32		0.72		Status: Ongoing	4.71	***	−1.92	***
Shipley: Abstraction	−0.07	***	0.04	*	Type: Health	−0.25		0.29	
Shipley: Vocabulary	−0.02		0.05	**	Type: Work	−1.54	***	0.97	***
Personality: Openness	0.06		−0.34	**	Type: School	−0.10		0.26	
Personality: Conscientiousness	−0.06		−0.07	*	Type: Local	−1.08	*	−0.17	
Personality: Extraversion	0.22	*	−0.03		Type: Financial	−2.88	***	1.35	***
Personality: Agreeableness	0.92	***	0.08						
Personality: Neuroticism	0.26	*	−0.03		**Baseline Attribute**	**Coeff.**	**p**	**Coeff.**	**p**
Positive Affect	−0.00		−0.03	**	SM: Num. Posts	1.28	***	-0.03	
Negative Affect	−0.00		0.00		SM: Avg. Post Length	8.62	*	-1.44	
Stai: Anxiety	0.02		−0.03	*	SR: Num. Records	-1.07	***	1.46	***
PSQI: Healthy Sleep Quality	−0.10	***	0.01		SR: Avg. Significance	0.50	***	-0.17	**

$Model_{3a}$: AIC = 920.3, Deg. Freedom= 34, LLk. = −425.15, χ^2= 2493.88, Pseudo R^2 = 0.75, p < 0.001 ***

$Model_{3b}$: AIC = 2160.4, Deg. Freedom= 34, LLk. = −1045.22, χ^2= 1468.90, Pseudo R^2 = 0.43, p < 0.001 ***