

# Observer Effect in Social Media Use

Koustuv Saha

University of Illinois  
Urbana-Champaign  
Urbana, IL, USA  
ksaha2@illinois.edu

Pranshu Gupta

Georgia Institute of Technology  
Atlanta, GA, USA  
prangupt@gatech.edu

Gloria Mark

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

Emre Kiciman

Microsoft Research  
Redmond, WA, USA  
emrek@microsoft.com

Munmun De Choudhury

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

## ABSTRACT

While social media data is a valuable source for inferring human behavior, its in-practice utility hinges on extraneous factors. Notable is the “observer effect,” where awareness of being monitored can alter people’s social media use. We present a causal-inference study to examine this phenomenon on the longitudinal Facebook use of 300+ participants who voluntarily shared their data spanning an average of 82 months before and 5 months after study enrollment. We measured deviation from participants’ expected social media use through time series analyses. Individuals with high cognitive ability and low neuroticism decreased posting immediately after enrollment, and those with high openness increased posting. The sharing of self-focused content decreased, while diverse topics emerged. We situate the findings within theories of self-presentation and self-consciousness. We discuss the implications of correcting observer effect in social media data-driven measurements, and how this phenomenon shines light on the ethics of these measurements.

## CCS CONCEPTS

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

## KEYWORDS

social media, observer effect, hawthorne effect, human behavior, self-presentation, language, causal-inference

## ACM Reference Format:

Koustuv Saha, Pranshu Gupta, Gloria Mark, Emre Kiciman, and Munmun De Choudhury. 2024. Observer Effect in Social Media Use. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3613904.3642078>

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

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0330-0/24/05...\$15.00

<https://doi.org/10.1145/3613904.3642078>

## 1 INTRODUCTION

The past decade has witnessed burgeoning research that has employed unobtrusively gathered social media data to infer various behavioral and psychological attributes and states of individuals [23, 34, 72]. Harnessing rapid advancements in machine learning, Facebook data, for instance, can allow us to identify an individual’s personality traits [97, 143], or assess if they are at risk of forthcoming mental illness [53]. Research claims a lot of promise in these pursuits—algorithms developed with social media data can support designing health interventions [43, 135], assisting decision-making in many contexts [74, 96], and providing actionable insights that have been difficult to gather through conventional social science methods that use self-reported information alone [35, 82, 138].

Most of the above research relies on retrospectively collected social media data—data created by subjects unaware of the possibility of it being used for algorithmic inferences. For social media data-driven algorithms to be usable and useful in the real world, these algorithms would have to go beyond showcasing feasibility on retrospective data to functioning accurately and reliably in prospective settings. However, different threads of recent research have argued that models trained on retrospective data do not necessarily translate well to the prospective setting due to bias and non-representativeness [25, 99, 130, 162]. For instance, Olteanu et al. argued that the validity and in-practice reliability of human-centered big data technologies suffer due to the unpredictability and complexity in human behavior along with unaccounted confounds [114]. Ruths and Pfeffer noted that studies harnessing social media data might misrepresent or be ineffective in the real world due to people’s changing behaviors [130]. Lazer et al. similarly unpacked how the Google Flu predictor algorithm, which used Google search data, overestimated the number of flu visits in real-time, despite performing exceptionally well on historical data [99].

In addition, privacy concerns may arise when people’s archival social media data is employed in making sensitive predictions without their consent. Fiesler and Proferes surveyed how Twitter users felt about their historical data being used for research without their knowledge or awareness and found that most respondents felt that researchers should not use postings without their consent [63]. Duffy and Chan found that social media users can alter their online self-presentation based on “imagined surveillance” on the platforms [57]. Scholars fear perceptions of surveillance when prospective research designs are adopted without participant awareness. For example, the Facebook emotional contagion study [98],

which did not seek consent from people whose Facebook feeds were modified for experimental purposes, was heavily critiqued on ethical grounds [90]. Pertinent here is the position of [boyd and Crawford](#), who noted that experiments conducted without participant awareness could reinforce the troubling perception of the technologies as “Big Brother, enabling invasions of privacy, decreased civil freedoms, and increased state and corporate control” [25].

An advocated solution to the issues of prospective research design is recruiting individuals through proper informed consent for data to be used in algorithms that infer behaviors and psychological states [62]. However, the prospective use of social media-based measurements poses new challenges yet to be addressed. Social media use is a form of intentional and conscious behavior or a behavior that individuals can alter at their will if they feel “observed”—changes that would be consistent with theories of social desirability, psychological reactance, self-presentation, and self-monitoring, to name a few [71, 152, 156, 156]. The *observer effect* is the phenomenon that individuals might deviate from typical behaviors, attributed to the awareness of being “watched” or studied [106, 115]. This phenomenon is also called the “research participation effect”, the “experimenter effect”, and the “Hawthorne effect” [40, 126].

The social-ecological model posits that human behavior is embedded in the complex interplay between an individual and their relationships, communities, and society [33]. While this theory explains the promise of social media as a viable source of naturalistic behavioral data, it points out a caveat—the observers (or researchers), who become a part of a subject’s ecology, may affect the subject’s behavior. Likewise, the ecological validity of these measurements remains unattested because the observer effect is not typically accounted for. The observer effect has been commonly cited to affect the reliability of observations in studies because it concerns research participation [94]. Consequently, [McCambridge et al.](#) noted, “If there is a Hawthorne effect, studies could be biased in ways that we do not understand well, with profound implications for research [87]”. Therefore, the observer effect remains a critical but unexplored phenomenon that may impact findings in social media research and our understanding of social media use.

Motivated by the above, in this paper, we broadly ask—**does observer effect present itself in prospective studies of social media, and if so, to what extent and how?** In particular, we study the following research questions:

**RQ1:** What is the prevalence and degree of the observer effect in social media use?

**RQ2:** How do individuals’ psychological traits explain their likelihood to show observer effect in social media behavior?

Given the lack of extant theoretical knowledge and empirical evidence of the prevalence and impact of the observer effect in social media use, our work is the first to operationalize and measure the observer effect in social media use. Our exploratory study aims to not only make theoretical and methodological contributions but also to spark our interest in measuring and accounting for this phenomenon in social media behaviors. We posit that quantifying the presence and degree of the observer effect can improve social media data-driven measurements. This would further provide clarity to researcher expectations and support developing measures to account

for this effect in study designs and findings in the computational social science field and its in-practice adaptations [130].

We conduct our investigation through a case study—a longitudinal, multi-disciplinary research effort—where 572 participants consented to social media (Facebook) data collection over a retrospective period of 82 months and a prospective period of 5 months from their enrollment date in the study. We operationalize the observer effect along two dimensions of social media use, comprising 266,320 Facebook postings, 1) behavioral changes, and 2) linguistic changes. Our analytic approach draws on two lines of research: first, causal inference methods [1] to minimize the impacts of confounding factors on changes in social media use, and second, modeling approaches in psychology that use clustering on psychological traits to derive person-centered changes. We employ time-series and statistical modeling to measure how participants deviated from their expected behaviors after enrolling in the above study, or in response to their awareness of being “observed”.

Our findings reveal that observer effect was indeed present, with posting behaviors of participants changing 17-34%, and linguistic attributes changing 4-57%. However, its occurrence varied across participants. For instance, individuals with high cognitive ability and low neuroticism showed an immediate decrease in social media posting after enrollment, but their behaviors got closer to typical behaviors over time. In contrast, individuals with high openness significantly increased posting quantity despite not showing any immediate posting changes following enrollment. Linguistically, most individuals decreased their use of first-person pronouns, which reflects reduced sharing of intimate and self-attentional content. While some individuals increased posting about public-facing events, others increased posting about social and family gatherings. We explain the behavioral changes with respect to psychological traits in a theory-driven fashion.

This study bears implications for methods that harness prospective social media data, and we discuss directions to account for the observer effect in social media study designs. Besides the measurement-related challenges induced by the observer effect, from an ethics point of view, this work empirically informs how interventions that leverage people’s digital data, can potentially interfere with their social media use. This can break the fundamental goals and expectations of social media platforms. Therefore, this work critiques the ethics of these measurements in the real world.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Observer Effect in Research

This effect has been commonly cited to affect the reliability of studies [94]. We adopt the following definition outlined in a systematic review by [McCambridge et al.](#) [106].

“The Hawthorne effect concerns research participation, the consequent awareness of being studied, and possible impact on behavior.”—[McCambridge et al.](#) [106]

Given that there are several arguments around the use and appropriateness of the term and context of “Hawthorne Effect” [40, 42, 106, 113], this paper adopts the term, “observer effect” for disambiguity and consistency purposes.

Several social science and psychology theories proposed in the last century explain behavioral change concerning the observer

effect in different settings. [Guerin](#) reviewed behavioral change in the presence of others (social presence), and postulated “social facilitation” and “social inhibition” as opposite effects on an individual’s performance. This effect is also described as a form of “reactivity” as individuals modify an aspect of behavior in *response* to a phenomenon (awareness of being observed) [93, 156]. Based on the social desirability theory, conformity and social desirability considerations can lead behavior to change in line with these expectations [42, 76]. The observer effect is also frequently studied in epidemiological and clinical studies to minimize confounds in findings [40, 68]. Further, this effect has also been attributed to affect methodologies such as field observations and ethnography [100], and is considered to be one of the biggest challenges and long described as the “Achilles heel” of participant research [115, 151].

Research has investigated limiting experimenter-observer interactions that may cause the observer effect [125]. Longitudinal studies have shown promises of mitigating such effects because participants either adapt to normalcy or become less aware of being observed over time [170]. From another perspective, this effect can be considered to be a strength (rather than a limitation) in certain settings, because it can potentially lead to more ethical, conscientious, and efficient behavior of participants [112], and greater inter-accountability due to co-presence [19].

However, none of the above may apply in our particular setting of social media sensing. Social media data, by its very nature, is naturalistically created by an individual through their self-motivated and self-initiated will and is collected passively and unobtrusively. An individual who consents to sharing social media data may not actively feel aware of being observed. This awareness might influence certain behavioral amendments that essentially normalize over time or a process known as *habituation* in behavioral sciences [39]. To understand the likelihood and extent of observer effect on social media behavior, we examine social media behavior following enrollment in a longitudinal study.

## 2.2 Behavior Change on Social Media

A vast body of research has studied social media behavior in a variety of ways, spanning prediction and inference studies on information dissemination, political interests, stock market, emotion, and health and wellbeing [3, 23, 34]. The growing evidence of the relationship between human behavior, psychology, and language allows us to infer these behavioral changes when we analyze longitudinal social media data. Like our physical world, people’s online presentation is a factor of their social network [57, 86]. [Guillory and Hancock](#) found that the public-facing nature of social media platforms such as LinkedIn influences an individual’s accountability and reduces deception, which also aligns with Donath’s early research on identity and deception in online spaces [56]. [Reinecke and Trepte](#) found that social media provides an environment for online authenticity, and authentic self-presentation contributes to positive psychological wellbeing [121]. Similarly, a body of literature reveals evidence regarding how social media facilitates candid self-disclosure for an individual [52, 124].

Prior work in HCI, social computing, and computational social science has studied behavioral changes on social media in several contexts. [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 [51]. [Golder and Macy](#) studied the variability in mood and sentiment on weekends and weekdays. Other longitudinal studies have examined behavioral changes around exogenous or endogenous, anticipated or unanticipated events, e.g., antidepressant use [136], counseling advisories [137], alcohol and substance use [95, 101], diagnosis with mental health conditions [60, 80, 81], suicidal ideation [54, 169], and so on. Relatedly, [Ernala et al.](#) adopted the Social Penetration Theory to operationalize intimacy of self-disclosure and audience engagement on social media [59], and [Ma et al.](#) conducted an online experiment to study the relationship between content intimacy, self-disclosure, and audience type on social media [102].

Researchers have also explored behavioral changes around topics related to observer effect, such as privacy. Back in 2014, when [Zhang et al.](#) studied “creepiness” and privacy concerns related to social media use, they found concerns shifting from boundary regulation to behavior tracking by social media platforms for targeted advertising [171]. However, social media- and web-based behavioral inferences have evolved since then and have also come under ethical and political scrutiny for privacy breaches such as the Cambridge Analytica scandal on Facebook [32]. This has also renewed attention to the challenges that may arise when data is appropriated for surveillance by different stakeholders, e.g., workplace surveillance [70, 92]. At the same time, concerns related to audience, boundary, and disclosure regulations are evident on social media; people want themselves to be viewed in particular ways across different audiences [57, 91, 104, 161]. As per Goffman’s theory of self-presentation, individuals may present two kinds of information (including on social media)—one that they intend to “give off” and one that “leaks through” without any intention [71, 111, 167]. One strategy of boundary regulation that is known to be prevalent on social media is self-censorship [49, 104]. Self-censorship occurs when social media users prevent themselves from posting or conducting a behavior despite a self-initiated initial desire to do so [49]. For example, [Wang et al.](#) studied self-censorship with respect to regretful thoughts [164]. Also, privacy concerns may lead to changes in social media behavior regarding presentation, censorship, and information sharing [4, 161]. However, researchers have also found an apparent “privacy paradox”, i.e., despite the awareness of privacy concerns, individuals may share more personal information on social media [15]. This shows that people’s social media behavior is complex and depends on each individual’s personality, perceptions, beliefs, and external factors [78, 120].

In addition, prior work has also revealed various factors that may describe why and how an individual self-describes themselves on a social media platform [31, 83, 102, 134]. While social media data is a useful signal to analyze behavioral changes, people’s perceptions about the use of social media may significantly affect their behavior [102, 134]. The current study examines this phenomenon by leveraging longitudinal social media data to delineate the impact of observer effect on people’s social media behavior.

### 2.3 Theories of Behavior Change in Research

Social scientists and psychologists have proposed numerous theories related to behavioral change. The socio-cognitive theory adopts an agentic perspective to human development, adaption, and change by distinguishing three modes of agency: personal, proxy, and collective [14]. The situated identity theory states that relevant cues in behavioral settings are first translated to identity potentials, which provide the basis for specific behavioral choices [8]. Self-consciousness is another construct that may influence one’s strategic self-presentation [11]. Again, the concept of psychological reactance describes that individuals have certain freedoms regarding behaviors, which, if reduced or threatened, they react to regain them [28]. Introduced by Snyder, the concept of self-monitoring posits that people self-monitor their self-presentations, expressive behavior, and non-verbal affective displays [152]. Self-monitoring can be considered a form of a personality trait that regulates behavior to accommodate social situations [152].

Further, Fishbein and Cappella noted, “Although there are many theories of behavioral prediction such as the Theory of Planned Behavior [6, 7], the Theory of Subjective Culture and Interpersonal Relations [160], the Transtheoretical Model of Behavior Change [118], the Information/ Motivation/ Behavioral-skills model [66], the Health Belief Model [20, 127, 128], Social Cognitive Theory [13, 14], and the Theory of Reasoned Action [64], a careful consideration of these theories suggest that there are only a limited number of variables that must be considered in predicting and understanding any given behavior [65]” [65]. They published an integrative model bringing together several theoretical perspectives [65].

This paper draws on the above theories and variables to interpret and explain the observer effect in social media behaviors. After quantifying the deviation in actual and expected post-enrollment behaviors, we investigate how people’s psychological traits could likely explain the changes by situating in the above theories.

## 3 STUDY AND DATA

The data for this study comes from the Tesseract project [105]. The study enrollment was conducted from January 2018 through July 2018. Participants either received a series of staggered stipends totaling USD \$750 or participated in a set of weekly lottery drawings (multiples of USD \$250 drawings) depending on their employer restrictions. At the time of enrollment, the participants responded to survey questionnaires related to demographics, and trait-based measures relating to personality, affect, sleep, and executive functions. The participants were requested to remain in the study for either up to a year or through April 2019. Participants were from various parts of the U.S., including from states in the north-western U.S. (Washington, Oregon, and Idaho), western and south-western U.S. (California, Utah, Arizona), central and eastern U.S. (Colorado, Minnesota, Iowa, Kansas, Missouri, Illinois, and Indiana), southern U.S. (Texas, Georgia, Alabama, and Tennessee), south-eastern U.S. (North Carolina), and north-eastern U.S. (New York, Washington DC, Pennsylvania, Massachusetts, West Virginia, Vermont), etc.

The Tesseract project focused on studying U.S. *information workers*—workers who process and work with information rather than physical objects, and in recent years, do so typically using computing technologies [103, 105]. Recruitment was done through workplace emails,

**Table 1: Summary of pre- and post- enrollment Facebook datasets of the participants.**

Type	Pre-Enrollment		Post-Enrollment	
	Range	Mean	Range	Mean
Posts	26-4,472	865	8-964	101
Comments	34-10,228	1,593	5-1,104	175
Likes	62-52,139	6,536	15-4,540	940
Duration (months)	0-160.27	82.52	0-12.87	4.59

messaging boards, and newspaper advertisements. Individuals were provided with an interest form via a Google Form, following which they were emailed a detailed consent form. Then, participants were enrolled through in-person and remote enrollment in early 2018. The in-person recruitments included researchers in the project team doing multiple rounds of corporate company site-visits to speak about and recruit participants. The remote enrollments were conducted via Zoom. The participant onboardings included explaining the study protocol and consenting process, and clarifying participant questions through researcher proctoring sessions. This was followed by participants responding to the survey questions. More details about the participant recruitment, study protocol, and challenges in setting up the study is in Mattingly et al. [105].

### 3.1 Social Media Data



















The Tesseract project asked consented participants to authorize their Facebook data, *unless they opted out or did not already use Facebook*. The participants authorized access to social media data through an Open Authentication (OAuth) based data collection infrastructure developed in Saha et al. [131]. OAuth protocol is an open standard for access delegation, commonly used for internet users to log in and grant third-party access to their information without sharing passwords. OAuth 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 [75] and its longitudinal nature has enabled several studies of human behavior [48, 53, 166], it suits our problem setting of understanding observer effect in social media behavior. Out of the total 572 participants who provided access to Facebook data, 532 made at least one post on their Facebook timeline. Table 1 summarizes the Facebook dataset of Tesseract participants, and we find that there are roughly 82 months data per participant in the pre-enrollment period and roughly 5 months data per participant in the post-enrollment period. For the scope of this study, we apply a filter of participants with at least 60 days of post-enrollment data to ensure sufficient data for examining observer effect—this leads to 316 participants, whom we study in the rest of this paper. Later in Section 6, we conduct robustness tests and repeat our experiments for other thresholds of the number of days to ensure our study design choices did not introduce biases in our findings.

### 3.2 Self-Reported Survey Data

Tesseract project’s enrollment process included initial demographics surveys (age, gender, education, income, type of occupation, role in the company, and income), and surveys of self-reported

**Table 2: Summary of demographics and individual differences of 316 participants whose data is studied for observer effect.**

Covariates	Value Type	Values / Distribution	
<i>Demographic Characteristics</i>			
Gender	Categorical	Male (171)   Female (145)	
Age	Continuous	Range (21:63), Mean = 36.36, Std. = 10.28	
Education Level	Ordinal	5 values [HS., College, Grad., Master's, Doctoral]	
Born in U.S.	Binary	Yes (283)   No (33)	
<i>Job-Related Characteristics</i>			
Income	Ordinal	7 values [<\$25K, \$25-50K, ..., >150K]	
Tenure	Ordinal	10 values [<1 Y, 1Y, 2Y, ..., 8Y, >8Y]	
Supervisory Role	Boolean	Non-Supervisor   Supervisor	
<i>Personality Trait (BFI)</i>			
Extraversion	Continuous	Range (1.7:5.0), Mean = 3.43, Std. = 0.71	
Agreeableness	Continuous	Range (2.3:5.0), Mean = 3.97, Std. = 0.57	
Conscientiousness	Continuous	Range (1.9:5.0), Mean = 3.90, Std. = 0.65	
Neuroticism	Continuous	Range (1.0:4.6), Mean = 2.52, Std. = 0.82	
Openness	Continuous	Range (2.2:5.0), Mean = 3.88, Std. = 0.59	
<i>Cognitive Ability (Shipley)</i>			
Fluid Cog. Ability (Abstraction)	Continuous	Range (5:24), Mean = 16.53, Std. = 3.32	
Crystallized Cog. Ability (Vocabulary)	Continuous	Range (18:40), Mean = 33.82, Std. = 3.63	
<i>Affect and Wellbeing</i>			
Positive Affect	Continuous	Range (13.0:49.0), Mean = 33.91, Std. = 5.84	
Negative Affect	Continuous	Range (10.0:40.0), Mean = 17.14, Std. = 5.24	
Anxiety	Continuous	Range (20.0:67.0), Mean = 39.01, Std. = 10.00	
Sleep Quality	Continuous	Range (1.0:16.0), Mean = 7.14, Std. = 2.75	

psychological constructs. 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 the study period, participants received daily or periodic validated surveys that recorded their self-reported assessments of job performance.

The collected psychological traits of individuals included 1) *Cognitive Ability* (or executive function), as assessed by the Shipley scales of Abstraction (fluid intelligence) and vocabulary (crystallized intelligence) [150], 2) *Personality Traits*, the big-five personality traits as assessed by the Big Five Inventory (BFI-2) scale [153, 159], and 3) *Affect and Wellbeing*, the general positive and negative affect levels as assessed through the Positive And Negative Affect (PANAS-X) scale [165], the anxiety level as measured via State Trait Anxiety Inventory (STAI-Trait scale) [154], and the quality of sleep as measured via the Pittsburg Sleep Quality Index (PSQI) scale [55]. Table 2 shows a descriptive summary of the 316 participants whom we study for observer effect.

### 3.3 Statistical Power of Participant Pool

While we cannot claim absolute representativeness of the U.S. information workforce, we see a diversity of participants across demographic and psychological traits (Table 2). Power analysis in statistics estimates the minimum sample size for a study to make significant inferences on a given population [157]. Likewise, we use power analysis to assess if this study has a sufficient sample size of participants to make reasonable inferences about the population. This study's participant pool belongs to information workers in the U.S. According to the U.S. Census Bureau, a rough estimate of the number of information workers in the U.S. is 4.6 million [122]. We calculate a sample size representative of this population with a 95%

confidence interval and 5% margin of error; this comes out to be a sample size of 385. Given that the net social media sample size is 574 participants, out of which usable data for studying observer effect is for 316 participants, this study assumes to have a reasonable sample of information workforce in the U.S.

### 3.4 Preliminary Analyses

**3.4.1 Quantity of Posting.** Posting behavior is a prominent social media behavior that has revealed significant signals of human behavior in prior work [51, 60]. We measure the average posting behavior of participants over time and around enrollment in the study. Figure 1a shows the daily average posting behavior of the participants relative to the day of enrollment, where day=0 corresponds to the enrollment day for the participants. We notice an apparent bump in the study's average number of posts per day post-enrollment.

**3.4.2 Expressive Behavior.** We examine the changes in the expressive behavior of the participants by using the psycholinguistic lexicon, Linguistic Inquiry and Word Count (LIWC) [158]. We obtain the psycholinguistic changes in the participants' posts after enrollment in the study. Figure 1b reports the effect sizes comparing pre- and post-enrollment normalized use of psycholinguistic categories across the participants. A positive effect size (Cohen's  $d$ ) indicates greater use of the category post-enrollment, whereas a negative effect size indicates lower use in the post-enrollment period. Cohen's  $d$  is considered to be a significant difference for magnitudes greater than 0.15. At an aggregate level, several psycholinguistic categories show significant changes. For example, within pronouns, *first-person pronoun* use decreases, which is associated with a decreased sharing of intimate content and decreased self-attentional

focus [45]. In contrast, the use of *first-person plural*, *second-person*, and *third-person pronouns* increases. We also find a decrease in the use of cognition-related words (e.g., *cognitive mechanics*, *discrepancies*, *inhibition*, *negation*, etc.). We also find a significant decrease in affective categories of *anger*, *sadness*, and *swear*.

The above preliminary analyses indicate certain changes in people’s behavioral and expressive social media use following enrollment in the study at an aggregated level. This motivates us to examine the changes in a much more rigorous and robust fashion. Given that not all individuals are the same, this study borrows from person-centric approaches to examine the changes in cohorts (clusters) of similar individuals on psychological constructs [50].

### 3.5 Privacy, Ethics, and Disclosure

The Tesserae project was approved by the Institutional Review Boards (IRBs) at the involved research institutions. The participants signed informed consent to provide their data for the study. The enrollment briefing and consent process explicitly explained that the study participation did not necessitate them to use social media in a particular fashion, and they were expected to continue their typical social media use. Our work is committed to securing the privacy of the participants. This paper uses de-identified data for analyses, conducted on secure encrypted servers, and provides paraphrased quotes to reduce traceability yet provide context in readership. We remove any information related to personal identity and paraphrase all quotations in this paper to avoid traceability. This paper, by design, does not use demographic data for clustering participants given the ethical concerns surrounding privacy intrusiveness, non-inclusivity, and discriminatory aspects of such approaches and their misuse [88]. Our research team comprises researchers holding diverse gender, racial, and cultural backgrounds, including people of color and immigrants, and hold interdisciplinary research expertise in the areas of HCI, UbiComp, machine learning, and psychology.

## 4 METHODS

### 4.1 Measuring the Observer Effect in Social Media Use

Theoretically, observer effect is a change in behavior because of being “observed” [106]. However, there are no established means to operationalize the observer effect, particularly in the context of social media use<sup>1</sup>. Our study, by design, considers enrollment for study participation as the *treatment*, and therefore, does not include a comparison/control group as enrolling this group would have subjected them to the same treatment and likely introduced biases of measuring the observer effect. Instead, we draw on synthetic control-based causal approaches [1, 2] that address the comparison group’s unavailability limitation by *synthetically* preparing control data through data-driven means.

We measure if enrollment in the study (treatment) *caused* the participants to change their social media use to their otherwise typical social media use. Social media use includes posting behaviors, engagements (likes and comments) received, and language use on the social media platform. We employ time series modeling

to predict participants’ *expected* post-enrollment social media use or the counterfactual data had they not enrolled in the study. We operationalize the observer effect as the post-enrollment deviation in the participants’ *observed* social media use from *expected* use. That is, we operationalize the observer effect ( $\alpha$ ) for a participant  $i$  and time period  $T$ , as the difference between their observed ( $Y^o$ ) and expected ( $Y^e$ ) social media use:  $\alpha_i[T] = Y_i^o[T] - Y_i^e[T]$ .

Social media data is prone to high variance across individuals, limiting the reliability of (participant) population-level analysis of the observer effect. Again, social media data is characteristically sparse, so it is challenging to extrapolate from individual behaviors [132]. We adopt a middle ground between fully generalized and fully personalized approaches by clustering individuals on self-reported psychological traits and conducting cluster-wise examinations of deviation in social media use. We measure the deviation in social media use along the dimensions of (1) behavioral changes: posts made and engagement received; and (2) linguistic changes: topics and psycholinguistics. This section describes our approach to clustering individuals (Section 4.2) and measuring observer effect per cluster (Section 4.3, 4.4).

### 4.2 Clustering Participants on Traits

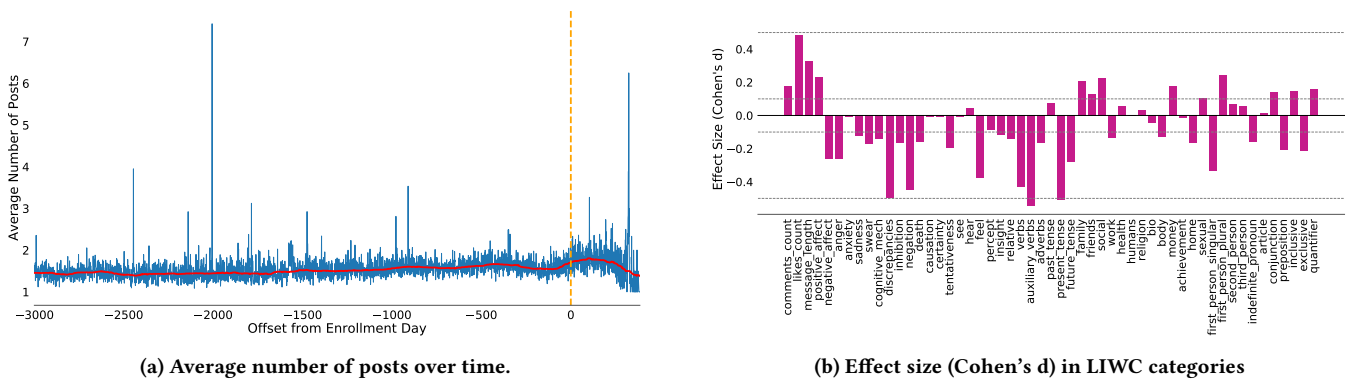
Although personalized approaches are the best means to study individual-level behavioral changes, it is hard to conduct personalized examinations on social media data because of sparsity issues that compromise statistical power. On the other hand, examining behaviors on the entire dataset (or variable-centered approaches) would suffer from high variance across individuals’ social media use (social media use can significantly vary across individuals). Therefore, drawing on prior work [132], we balance the trade-offs between too-personalized and too-generalized models by clustering individuals on self-reported psychological traits. Then, we measure the observer effect per cluster. This approach accounts for between-individual homogeneity and within-individual heterogeneity [132].

Given that demographic information are often privacy-intrusive, demographically discriminatory and non-inclusive [88], we conduct our clustering based on self-reported psychological traits of cognitive ability (abstraction and vocabulary) [150], Big-5 personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) [153], and affect and wellbeing (positive affect, negative affect, anxiety, and sleep quality) measures [55, 154, 165]. Using these psychological traits as features, we conduct  $k$ -means clustering on the individuals.

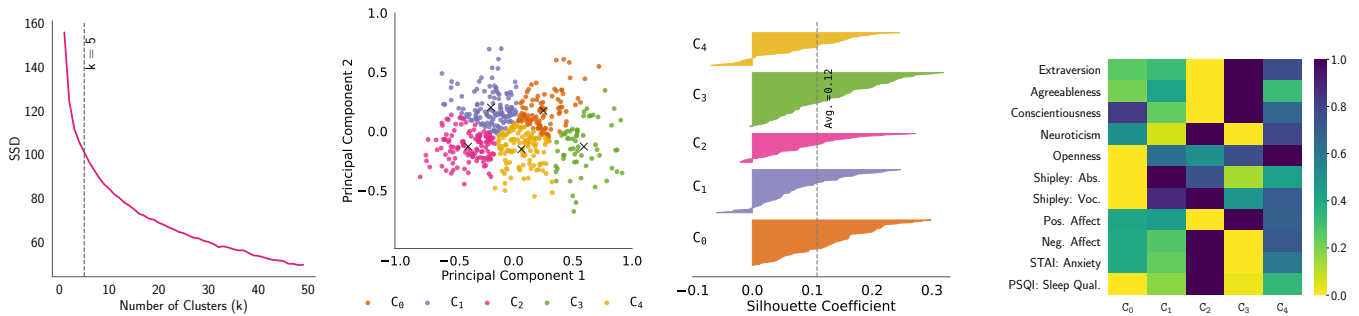
We employ the elbow heuristic to obtain the optimal number of clusters ( $k$ ) in our approach [140]. Figure 2 shows the elbow plot of the mean sum of squared distances to the cluster centroids and the number of clusters ( $k$ ), roughly estimating an optimal number of clusters at  $k=5$ . This leads us to cluster the initial 532 individuals in the dataset into five clusters ( $C_0$  to  $C_4$ ), containing 98, 121, 92, 146, and 83 members, respectively. Figure 3 shows a scatter-plot visualization of the clusters and their centroids in a two-dimensional Principal Component Analysis (PCA)-reduced space.

**4.2.1 Evaluating Cluster Heterogeneity.** We evaluate if our clustering approach actually reduces the heterogeneity in data per cluster. Table 3 shows a comparison of the standard deviation of traits

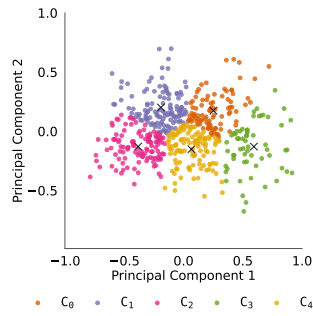
<sup>1</sup>Note that this paper uses “social media use” as a phrase encompassing social media posting behaviors, engagements received, and language use.



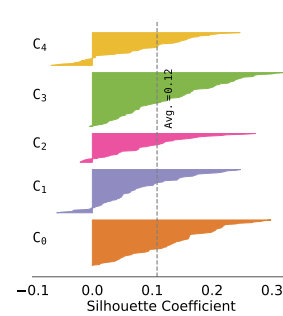
**Figure 1: (a) Average number of posts per day across all participants around relative offset from the day of enrollment. Day 0 indicates the enrollment day. (b) Effect size (Cohen's *d*) comparing before and after enrollment datasets of users across psycholinguistic (LIWC [158]) attributes. A positive Cohen's *d* indicates the use increased in the post-enrollment period, and a negative Cohen's *d* indicates the use decreased in the post-enrollment period.**



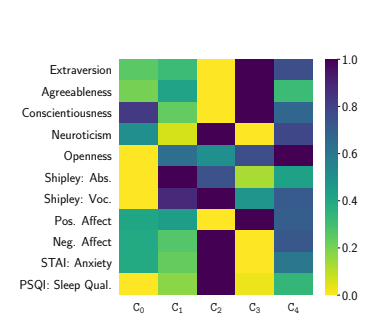
**Figure 2: Elbow plot of the number of clusters (*k*) and the mean sum of squared distances to centroids (SSE).**



**Figure 3: PCA-reduced plot visualizing the clusters in two dimensions. The "x"s are the KMeans cluster centroids.**



**Figure 4: Silhouette plot for Kmeans clustering (*k*=5). The cluster widths represent the number of datapoints in it.**



**Figure 5: Distribution of traits per participant cluster. The intensity of colors depicts the scaled mean of the trait.**

**Table 3: Comparison of the mean and standard deviations (std.) in traits in the entire data and that per cluster, one-way ANOVA (*F*-statistic), statistical significance reported as *p*-value, \*<0.05, \*\*<0.01, \*\*\*<0.001.**

Trait	All		C <sub>0</sub>		C <sub>1</sub>		C <sub>2</sub>		C <sub>3</sub>		C <sub>4</sub>		<i>F</i> -stat.
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	
<b>Personality Traits</b>													
Extraversion	3.43	0.71	3.14	0.68	3.20	0.56	2.88	0.61	3.88	0.50	3.64	0.62	22.23***
Agreeableness	3.97	0.57	3.70	0.42	3.86	0.52	3.55	0.52	4.29	0.41	3.78	0.55	15.94***
Conscientiousness	3.90	0.65	4.16	0.53	3.42	0.48	3.12	0.52	4.37	0.39	3.96	0.44	46.67***
Neuroticism	2.52	0.82	2.64	0.58	1.97	0.39	3.39	0.46	1.87	0.48	3.07	0.58	78.74***
Openness	3.88	0.59	3.13	0.48	3.88	0.42	3.72	0.43	40.02	0.52	4.30	0.35	25.85***
<b>Cognitive Ability</b>													
Shipley: Abs.	16.53	3.32	16.49	3.02	17.93	2.84	17.57	3.32	16.68	2.83	17.11	3.15	2.95**
Shipley: Voc.	33.82	3.63	31.90	3.62	34.05	2.85	34.33	3.44	33.07	3.50	33.64	3.60	1.70*
<b>Affect and Wellbeing</b>													
Pos. Affect	33.91	5.84	32.49	4.51	32.86	4.78	28.33	4.88	38.51	3.91	35.52	5.11	25.23***
Neg. Affect	17.14	5.24	17.33	3.35	16.24	3.70	22.49	4.32	13.95	2.55	20.16	4.63	23.09***
STAI: Anxiety	39.01	10.00	38.63	5.29	35.21	5.08	51.81	7.01	30.11	5.32	43.24	7.64	79.28***
PSQI: Sleep Qual.	7.14	2.75	6.00	2.29	6.41	2.58	8.37	2.92	6.07	2.59	6.80	2.08	9.12***

in the entire data against that per cluster, and one-way ANOVA (*F*-statistic). We find that the standard deviation of each trait per cluster is lower than that in the entire data. One-way ANOVA essentially measures the ratio of between-group variance and within-group

variance—a *F*-statistic more than 1 indicates between-cluster variance is greater than within-cluster variance. Therefore, Table 3 suggests that our clustering sufficiently distinguishes the clusters on these traits with statistical significance. Also, the silhouette plot

**Table 4: Descriptions of clusters (C) on traits. Note PSQI magnitude has a reverse interpretation with sleep quality (e.g., higher PSQI indicates lower sleep quality) [55].**

C	N	Trait Overview	Persona Characteristics
C <sub>0</sub>	60	High (Conscientiousness, Sleep Quality), Low (Openness, Cognitive Ability)	Routine-oriented
C <sub>1</sub>	66	High (Cognitive Ability), Low (Neuroticism)	Emotionally-stable and innovative
C <sub>2</sub>	44	Low (Extraversion, Agreeableness, Conscientiousness, Pos. Affect, Sleep Quality), High (Neuroticism, Cognitive Ability, Neg. Affect, Anxiety)	Withdrawn and prone to stress and irritability
C <sub>3</sub>	97	High (Extraversion, Agreeableness, Conscientiousness, Pos. Affect, Sleep Quality), Low (Neuroticism, Neg. Affect, Anxiety)	Positive, friendly, and well-balanced
C <sub>4</sub>	49	High Openness	Curious and adventurous

(Figure 4) shows that each of the clusters has a significant number of data points above the average silhouette score, and there are no wide fluctuations in the silhouette sizes—ensuring that the clustering did not yield sub-optimal clusters [50, 129]. These examinations reveal cluster validity [168] in our approach.

#### 4.2.2 Characterizing and Describing the Clusters of Individuals.

Fig. 5 shows the average distribution of the traits and Table 4 summarizes the characteristics of the five clusters. We draw on the literature [16, 46] to assign persona characterization for these clusters, which we describe below:

—Cluster C<sub>0</sub> has individuals with high conscientiousness and sleep quality and low openness and cognitive ability, suggesting that they are likely to be *routine-oriented* [16].

—Cluster C<sub>1</sub> has individuals with high cognitive ability and low neuroticism, so they are more likely to be *emotionally stable and innovative* [16, 150].

—Cluster C<sub>2</sub> has individuals with high neuroticism, cognitive ability, negative affect, and anxiety, and low extraversion, agreeableness, conscientiousness, positive affect, and sleep quality. These characteristics suggest that they are likely to be more *withdrawn, disagreeable, and prone to stress and irritability* [16]. The ARC taxonomy describes this cluster of individuals as “overcontrolled”, who would likely show obsessive-compulsive and avoidant symptoms [46].

—Cluster C<sub>3</sub> has individuals with high extraversion, agreeableness, conscientiousness, positive affect, and sleep quality, but low neuroticism, negative affect, and anxiety. They can be characterized to be *positive, friendly, and well-balanced*, i.e., resistant and less likely to experience stress, anxiety, and negative emotions. The ARC taxonomy describes their combination of personality traits as “resilient”, and they likely show high psychological adjustments [46].

—Cluster C<sub>4</sub> has individuals with high openness. Those with high openness tend to be *curious and adventurous*—more open-minded and willing to embrace fresh ideas and novel experiences [41].

### 4.3 Measuring Behavioral Changes

**4.3.1 Measures to Quantify Behavioral Changes.** We examine the post-enrollment changes in posting and engagement:

– *Posting Behavior.* We examine the social media posting behavior of participants in terms of the daily average number of 1) posts (*quantity of posts*) and 2) words (*verbosity of posting*).

– *Engagement Received.* We examine the engagements received by the participants on their social media posts in terms of the daily average number of 1) *likes* and 2) *comments* received.

**4.3.2 Modeling and Quantifying Behavioral Changes.** Drawing on interrupted time series and synthetic control-based causal approaches [18, 107], we compute the deviation in actual behavior from the expected behavior of the participants as modeled on their historical behavior. For each cluster, drawing on prior work [138], we build autoregressive models to extrapolate post-enrollment expected behaviors of the participants. We build the models accounting for trends and seasonalities in the time series. We train the models on the pre-enrollment data, using an 80:20 split (80% for training and 20% held-out for testing), and apply grid search to optimize for the best parameters of the time series prediction models. We conduct parameter tuning on the training dataset by dividing it into 80%-20% validation set. In particular, we use the Seasonal Autoregressive Integrated Moving Average Exogeneous (SARIMAX) model from the STATSMODEL library [69, 144], which takes parameters of order ( $p, d, q$ ) and seasonal order ( $P, D, Q, s$ ). For every cluster, we iterate on all possible combinations of parameters to build models, which are then decomposed on trends and seasonalities to predict the validation sets. By adopting symmetric mean absolute percentage error (SMAPE) and alkaline information criterion (AIC) as our evaluation metrics, we dynamically obtain the parameter combinations for best-performing models per cluster. We further evaluate the models on the 20% held-out data through SMAPE, which quantifies errors in the range of 0 to 100, where lower values indicate a better predictive model. We study the differences between observed and expected behaviors in the short-term (two-weeks) and long-term (100-days) post-enrollment period and measure the statistical significance of the differences using paired  $t$ -tests and effect size (Cohen’s  $d$ ). We also compute the slope changes in the time series of social media use from pre- to post-enrollment periods and draw on Brodersen et al. [29] to measure causal impact (CI). Higher values of the posterior probability of CI would indicate a significant behavioral change after enrollment in the study. We conduct the causal impact analysis using the CAUSALIMPACT library [30].

### 4.4 Measuring Linguistic Changes

We examine the changes in linguistic expressiveness on social media posts, through two analyses, 1) topics and 2) psycholinguistics. For both analyses, we compare the pre-enrollment and post-enrollment data of the individuals. We describe the analyses below:

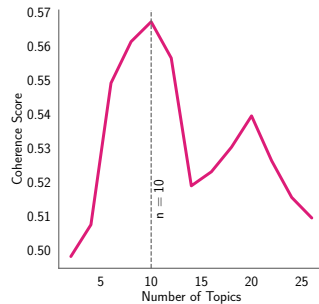
**4.4.1 Measuring Topical Changes.** Topics are useful for understanding the content of people’s social media expressions [36]. We conduct topic modeling in our dataset to examine how the prevalence and diversity of topics evolve following study enrollment. To extract topics automatically, we employ the widely adopted Latent Dirichlet Analysis (LDA) on the dataset [22]. LDA is known to produce stable and interpretable topics and has often been used in social media and human behavior research [36, 60, 123].



**Table 5: Thematic categories of identified topics and example paraphrased posts (italicized) in our dataset.**

Theme	Topic Words
Travel & Locations	country, green, baby, miss, right, chicago, sad, need, let, denver, mean, airport, hello, way, win, begin, yum, national, cubs, joanie <i>Smiles all around after a good ATD conference together in Denver.</i> <i>Annoyed to have missed our flight to Belgium after two hours delay on a plane from Dallas to London. Stuck in the airport for 8 hours.</i>
Food & Drinks	lol, new, ready, room, sweet, boy, getting, waiting, finally, time, chicken, need, delicious, chicken, got, cheese, food, beer, gotta, yeah, guess <i>Chicken on the grill, beef roast on the cutting board, regular and sweet potatoes in the oven. Guess who's not cooking tomorrow!</i> <i>Finally made it to a beer fest!</i>
Holiday Plans	christmas, school, vote, today, true, high, trip, look, season, awesome, johnson, merry, news, summer, party, check, mom, family <i>Morning hike, trip to the beach, and relaxing at our rental!</i> <i>It was such an amazing Christmas themed day for our family. My daughter would not sit with Santa alone, so we all did it together.</i>
News & Information	like, people, time, things, trump, think, watch, know, looks, right, got, thing, want, need, good, going, bad, stop, run, better, org <i>Climate models want to change the way we live ... should we listen? It's a short video, watch it.</i> <i>Trump's personal financial disclosure report is due Tuesday. Under the Ethics in Gov't Act, he has to disclose liabilities that exceeded \$10,000 in 2017.</i>
Work-Life Balance	home, work, day, got, yes, new, today, time, tomorrow, little, house, going, like, car, snow, hours, bed, dog, night, way <i>After work. Only one thing on my mind.</i> <i>Yay!!! No work for the whole next week!</i>
Family Gathering	good, morning, great, night, fun, day, time, weekend, dinner, week, friday, today, tonight, party, work, family, team, going, view, date, girls, weekend <i>Had a great visit with Otto &amp; family!</i> <i>Off to Los Angeles for an awesome family gathering to celebrate my grandma's 100th birthday.</i>
Social & Sports	game, want, tony, retweeted, play, south, come, bend, dame, notre, it's, tulio, tickets, world, need, free, shit, dace, wants <i>Watched my team in India play a friendly cricket match last night and got a lesson on the difference between batting in baseball versus cricket.</i> <i>Anyone looking for a couple of tickets to the Florida State game happening on Saturday?</i>
Greetings & Celebration	day, happy, love, birthday, wedding, today, anniversary, halloween, disney, beautiful, mom, http, best, little, year, life, wish, challenge, thank <i>Wishing my beautiful daughter a wonderful birthday. Love you baby girl.</i> <i>Made a rainbow cake to celebrate our visit with my best friend, Sonia!</i>
Friends & Family	years, time, love, family, friends, year, life, thanks, kids, amazing, best, know, today, old, wait, great, ago, days, help, people <i>Enjoying St Helena, brunch, and wine tasting with my son and friends.</i> <i>Thank you all for the kind wishes! It is so good to feel the love of friends from childhood all the way through to the present.</i>
Activities & Interests	like, read, years, wow, know, love, good, think, people, music, interesting, post, facebook, copy, wheels, place, favorite, book <i>First book I've read in a long time that I couldn't put down. The Life We Bury</i> <i>I made some delicious cookies with my favorite dudes tonight! Thanks for coming over and enjoying my hobby with me. You are all artists!</i>

**Finding Optimal Number of Topics.** To identify the optimal number of topics in our dataset, we draw recommendations from prior work [38, 73, 163] to vary the number of topics up to 25, and semi-automatically evaluate the quality of topic models, by combining the use of topical coherence scores as well as manual evaluations. Topical coherence score quantifies the degree of semantic similarity between high-scoring words within a topic [108]. Figure 6 plots the coherence scores on varying the number of topics from 2 to 26, suggesting that the highest coherence is achieved at around the number of topics ( $n$ ) as 10. The first author and two collaborators in the research team manually evaluated the topical distribution for  $n=8$ ,  $n=10$ , and  $n=12$  to find that the topical distributions at  $n=8$  and  $n=12$  were less semantically coherent, with a substantial increase in noisy keywords. Therefore, as guided by both coherence scores and manual examination, we use topic modeling for  $n=10$  topics for our study.



**Figure 6: Topical coherence scores on LDA topic modeling.**

**Interpreting Topics.** After building the topic models, we assign interpretable labels to topics and keywords. For this purpose, three members of the research team designed an interpretive annotation

to identify coherent themes in the keywords per topic. The topics were first inductively and independently coded with implied themes. Then the codes were compared and agreed upon to assign final thematic labels per topic. The thematic category of a topic was implied from the within-topic coherence and between-topic separation of keywords. These themes are 1) *Travel and Locations*, 2) *Food and Drinks*, 3) *Holiday Plans*, 4) *News and Information*, 5) *Work-Life Balance*, 6) *Family Gathering*, 7) *Social and Sports*, 8) *Greetings and Celebration*, 9) *Friends and Family*, and 10) *Activities and Interests*. Table 5 shows the 10 thematic categories and top occurring keywords per topic, along with example paraphrased post from our dataset.

**4.4.2 Measuring Psycholinguistic Changes.** Another dimension to understand people's expressiveness is through psycholinguistics [53, 143]. We used the psycholinguistically validated and widely adopted lexicon of Linguistic Inquiry and Word Count (LIWC) [158]. LIWC allows to categorize the pre- and post- enrollment social media data into psycholinguistic categories of: 1) *affect* (anger, anxiety, negative and positive affect, sadness, swear), 2) *cognition* (causation, inhibition, cognitive mechanics, discrepancies, negation, tentativeness), 3) *perception* (feel, hear, insight, see), 4) *interpersonal focus* (first person singular, second person plural, third person plural, indefinite pronoun), 5) *temporal references* (future tense, past tense, present tense), 6) *lexical density and awareness* (adverbs, verbs, article, exclusive, inclusive, preposition, quantifier), and 7) *personal and social concerns* (achievement, bio, body, death, health, sexual, home, money, religion, family, friends, humans, social).

## 5 RESULTS

This section describes our results—the first subsection shows deviations in social media behaviors and language use where we report two kinds of results, short-term (two-weeks period) and long-term (100-days period) deviation, and 2) in the second subsection, we validate and explain our findings with respect to psychological traits of individuals.

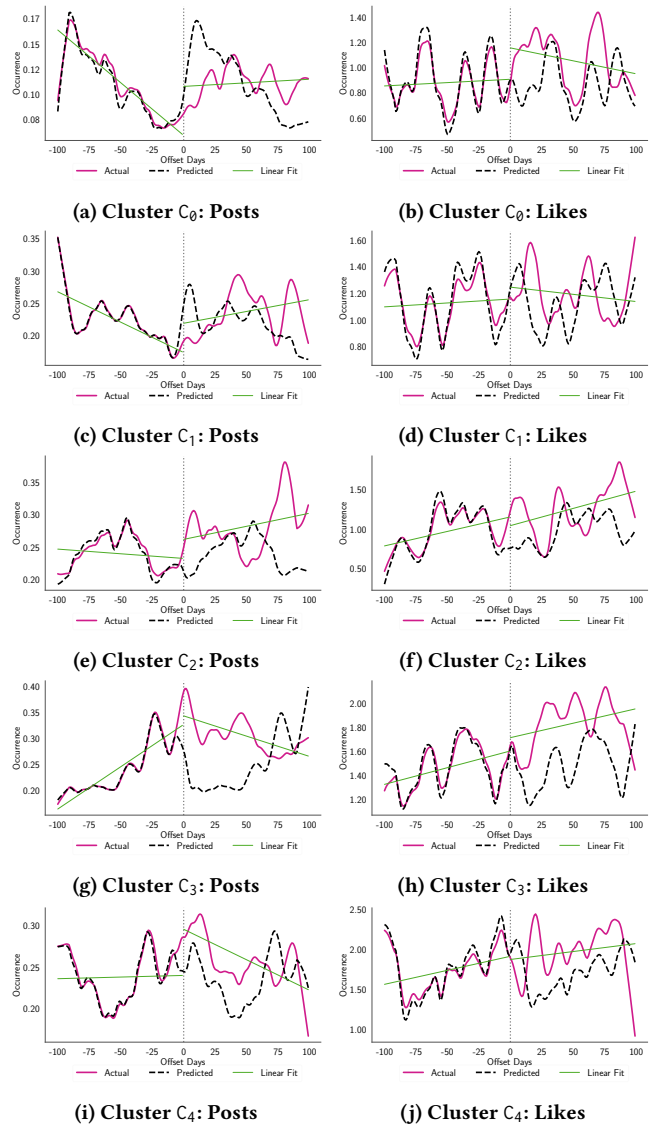
### 5.1 RQ1: Prevalence and Degree of Observer Effect in Social Media Use

Recall that we measure the deviation in social media use along the dimensions of (1) behavioral changes: posts made and engagement received; and (2) linguistic changes: topics and psycholinguistics. We obtain expected post-enrollment social media use by extrapolating pre-enrollment behavioral trends into the 100-day post-enrollment period through time series-based modeling (SARIMAX).

**5.1.1 Deviation in Behavior.** We extrapolate expected behaviors through ARIMA models using the pre-enrollment data, accounting for trends and seasonalities in time series, and measure the deviation in the actual post-enrollment measures from the expected measures. Table 6 summarizes the model metrics and changes in participants' social media use. Table 7 summarizes the slope changes in the time series of social media use from pre- to post-enrollment periods, along with causal impact computed as per Brodersen et al. [29]. High posterior probabilities of causal impacts (CI) indicate that the behaviors changed after enrollment in the study. Figure 7 show cluster-wise deviations in actual and expected time series of the number of posts and likes, which we describe below.

**Changes in Posting Behavior.** To obtain expected posting behaviors, the SARIMAX models predicting number of posts and words show mean symmetric mean absolute percentage errors (SMAPE) of 6.27 and 13.05, respectively. However, the deviation between predicted and actual values in the post-enrollment data is higher. In the 100-day post-enrollment data, clusters  $C_2$  and  $C_3$  show statistically significant deviations in both quantity and verbosity of posts, i.e., they posted significantly more frequently and longer than their expected behaviors— $C_2$  show an average 17% higher and  $C_3$  show an average 24% higher than the expected quantity of posts. Focusing on the initial two-weeks post-enrollment,  $C_2$  and  $C_3$  show similar (36% and 70%) increases in posting.  $C_0$  and  $C_1$  show respectively 44% and 26% lower frequency of posting in the first two weeks, but their posting behavior became closer to their typical posting behaviors after the initial two-week period. Interestingly, even though  $C_4$  seemed to post greater than expected, their posting behavior had a decreasing trend (negative slope).  $C_4$  individuals posted 41% shorter than expected posts in the initial two weeks.

**Changes in Engagement Received.** The ARIMA models predicting the expected number of comments and likes show mean SMAPEs of 20.76 and 15.15, respectively. In the 100-day post-enrollment period,  $C_3$  received an average of 25% and 22% higher than expected likes and comments respectively, and  $C_2$  received an average 29% higher than expected likes. As noted above, the received engagements are likely correlated to these individuals' higher posting activity.



**Figure 7: Evolution of the daily average number of posts and likes per cluster in 100-day pre- and post-enrollment periods. The dotted line in the center is the enrollment day (day 0).**

Considering two-week deviations, we find that  $C_2$ 's posts received an immediately higher quantity of comments (67%) and likes (96%), and  $C_4$  received 27% lower than expected comments.  $C_0$  and  $C_1$  did not have any significant deviations in the engagements received.

#### 5.1.2 Deviation in Language Use.

**Changes in Topical Themes.** Table 8 summarizes cluster-wise relative change in topical prevalence from pre- to post-enrollment.  $C_0$  individuals increased posting about public-facing topics, such as travel, food, and news, increased posting about family gatherings, but decreased posting about sports and celebratory events.  $C_1$  individuals increased posting about holiday plans, family gatherings,

**Table 6: Summary of behavioral deviations in post-enrollment compared to expected (or predicted) behavior per cluster in terms of SMAPE, paired *t*-tests, and effect size (Cohen's *d*). Statistical significance reported as *p*-value, \**p*<0.05, \*\**p*<0.01, \*\*\**p*<0.001. Positive *t* or *d* indicates higher values in the actual time series compared to the predicted time series. Significant values are shaded in blue to indicate an increase and red to indicate a decrease during the post-enrollment period.**

Cluster	Model	100-days post-enrollment					2-weeks post-enrollment				
		SMAPE	Mean (Act.)	Mean (Exp.)	SMAPE	t-test	Cohen's d	Mean (Act.)	Mean (Exp.)	SMAPE	t-test
<b>Posting Behavior</b>											
Average Daily Number of Posts											
Cluster C <sub>0</sub>	11.09	0.11	0.11	24.45	-0.05	-0.01	0.09	0.16	30.73	-4.31 ***	-1.59
Cluster C <sub>1</sub>	4.42	0.24	0.22	14.85	1.52	0.21	0.20	0.27	17.82	-3.68 ***	-1.35
Cluster C <sub>2</sub>	5.78	0.28	0.24	17.45	3.49 ***	0.49	0.30	0.22	19.77	3.93 ***	1.44
Cluster C <sub>3</sub>	4.20	0.31	0.25	18.00	5.76 ***	0.82	0.34	0.20	27.1	4.99 ***	1.84
Cluster C <sub>4</sub>	5.85	0.26	0.24	16.25	2.02 *	0.29	0.31	0.28	17.04	1.07	0.39
Average Daily Number of Words											
Cluster C <sub>0</sub>	22.69	0.67	0.59	42.85	1.10	0.16	0.58	0.75	54.22	-0.97	-0.36
Cluster C <sub>1</sub>	11.24	1.68	1.86	24.99	-1.44	-0.2	1.66	1.76	19.46	-0.38	-0.14
Cluster C <sub>2</sub>	11.31	2.14	1.80	24.5	2.60 *	0.37	2.24	1.72	23.28	1.46	0.54
Cluster C <sub>3</sub>	6.40	1.85	1.60	17.86	3.03 ***	0.43	1.84	1.40	18.15	1.96 *	0.72
Cluster C <sub>4</sub>	13.65	2.20	2.23	25.37	-0.18	-0.03	2.03	3.46	34.4	-3.72 ***	-1.37
<b>Engagement Received</b>											
Average Daily Number of Comments Received											
Cluster C <sub>0</sub>	39.29	0.16	0.13	51.71	1.16	0.16	0.18	0.18	56.14	-0.41	-0.15
Cluster C <sub>1</sub>	11.20	0.21	0.21	30.22	-0.35	-0.05	0.21	0.22	27.45	-0.41	-0.15
Cluster C <sub>2</sub>	18.61	0.26	0.25	33.23	0.26	0.04	0.35	0.21	32.51	2.57 *	0.94
Cluster C <sub>3</sub>	9.08	0.33	0.27	24.93	2.78 *	0.39	0.25	0.25	18.18	-0.18-	-0.07
Cluster C <sub>4</sub>	25.63	0.30	0.29	31.38	0.46	0.07	0.24	0.33	26.55	-2.13 *	-0.78
Average Daily Number of Likes Received											
Cluster C <sub>0</sub>	25.59	1.06	0.88	41.90	1.64	0.23	1.06	0.75	52.27	1.10	0.40
Cluster C <sub>1</sub>	10.74	1.20	1.13	28.27	0.68	0.10	1.13	1.35	29.33	-0.97	-0.36
Cluster C <sub>2</sub>	17.66	1.26	0.98	31.37	3.01 ***	0.43	1.49	0.76	33.87	3.10 ***	1.14
Cluster C <sub>3</sub>	8.04	1.84	1.47	18.43	4.74 ***	0.67	1.47	1.35	17.45	0.72	0.26
Cluster C <sub>4</sub>	13.74	1.97	1.73	28.2	1.70 *	0.24	1.57	1.94	32.34	-1.28	-0.47

**Table 7: Summary of behavior changes with causal impact (CI) estimation, with the slope in pre- and post- enrollment data, relative change in slope, Kolmogorov–Smirnov-test (KS-test), and posterior probability of causal impact (PP% CI) [29].**

	Pre-enrollment	Post-enrollment	Change %	KS	PP% CI		Pre-enrollment	Post-enrollment	Change %	KS	PP% CI
<b>Posting Behavior</b>						<b>Engagement Received</b>					
Average Daily Number of Posts						Average Daily Number of Comments Received					
C <sub>0</sub>	-1.05 × 10 <sup>-3</sup>	7.13 × 10 <sup>-5</sup>	106.80	0.47***	65.83	C <sub>0</sub>	-8.80 × 10 <sup>-5</sup>	-1.84 × 10 <sup>-4</sup>	-108.92	1.0***	58.94
C <sub>1</sub>	-9.39 × 10 <sup>-4</sup>	3.65 × 10 <sup>-4</sup>	138.92	0.48***	96.20	C <sub>1</sub>	-1.12 × 10 <sup>-4</sup>	1.72 × 10 <sup>-4</sup>	252.91	0.47***	53.35
C <sub>2</sub>	-1.44 × 10 <sup>-4</sup>	4.05 × 10 <sup>-4</sup>	380.74	1.0***	99.60	C <sub>2</sub>	3.42 × 10 <sup>-4</sup>	2.18 × 10 <sup>-4</sup>	-36.30	0.36***	99.10
C <sub>3</sub>	1.63 × 10 <sup>-3</sup>	-7.85 × 10 <sup>-4</sup>	-148.01	0.63***	100.00	C <sub>3</sub>	7.75 × 10 <sup>-4</sup>	5.50 × 10 <sup>-5</sup>	-92.90	1.0***	65.23
C <sub>4</sub>	3.93 × 10 <sup>-5</sup>	-7.29 × 10 <sup>-4</sup>	-1954.83	0.76***	91.11	C <sub>4</sub>	-2.22 × 10 <sup>-5</sup>	2.32 × 10 <sup>-4</sup>	1144.37	0.89***	91.71
Average Daily Number of Words						Average Daily Number of Likes Received					
C <sub>0</sub>	-3.60 × 10 <sup>-3</sup>	-4.19 × 10 <sup>-4</sup>	88.37	0.66***	82.82	C <sub>0</sub>	5.29 × 10 <sup>-4</sup>	-2.06 × 10 <sup>-3</sup>	-489.06	1.0***	75.62
C <sub>1</sub>	6.04 × 10 <sup>-5</sup>	-7.03 × 10 <sup>-3</sup>	-11740.23	0.73***	77.52	C <sub>1</sub>	5.86 × 10 <sup>-4</sup>	-1.08 × 10 <sup>-3</sup>	-284.34	0.84***	99.10
C <sub>2</sub>	2.20 × 10 <sup>-4</sup>	3.53 × 10 <sup>-3</sup>	1501.28	1.0***	98.40	C <sub>2</sub>	3.67 × 10 <sup>-3</sup>	4.36 × 10 <sup>-3</sup>	18.96	0.75***	99.80
C <sub>3</sub>	3.23 × 10 <sup>-3</sup>	-2.30 × 10 <sup>-3</sup>	-171.11	0.91***	96.30	C <sub>3</sub>	2.78 × 10 <sup>-3</sup>	2.41 × 10 <sup>-3</sup>	-13.33	1.0***	89.31
C <sub>4</sub>	-1.35 × 10 <sup>-2</sup>	2.08 × 10 <sup>-3</sup>	115.45	0.46***	52.35	C <sub>4</sub>	3.49 × 10 <sup>-3</sup>	1.98 × 10 <sup>-3</sup>	-43.22	0.89***	86.91

and celebratory events but decreased posting about news-related content. C<sub>2</sub> individuals showed the least changes in the expressiveness of content, and decreased posting about food and social events. C<sub>3</sub> individuals showed varied changes, with increased sharing about travel, food, and sports-related content, whereas a decrease in more personal content such as holiday plans, work-life balance, family, and celebratory events. Finally, C<sub>4</sub> individuals increased posting about food and family gatherings, whereas decreased posting about holiday plans, news, and interests-related content.

*Changes in Psycholinguistic Use.* Table 9 shows the changes in psycholinguistic use, which we examine below.

—Cluster C<sub>0</sub> individuals did not show any significant change in affective expressions except for *anger*. In cognitive expressions, these participants increased using words related to certainty. In perception, *feel* and *see* decreased, whereas *hear* increased. They also decreased *first person singular pronoun* use but increased *first person plural pronoun* use. We also find a decrease in several function words, including *adverbs*, *verbs*, *auxiliary verbs*, *quantifiers*,

**Table 8: Changes in topical prevalence post-enrollment in the study. Statistical significance is computed as per independent-sample  $t$ -tests (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). Significant values are shaded in blue for increased sharing, i.e., the higher average value in post-enrollment, and red for decreased sharing, i.e., the lower average value in the post-enrollment period.**

Topic	% Change in Cluster				
	C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
Travel & Locations	28.38 ***	-0.98	-7.69	25.14 *	-1.94
Food & Drinks	37.16 ***	2.28	-13.85 **	3.39 *	14.20 **
Holiday Plans	18.22 *	18.65 *	-7.22	-12.10 *	-10.21 *
News & Information	33.89 **	-14.25 ***	-6.19	-19.29 ***	-17.36 *
Work-Life Balance	-0.05	1.28	-8.93	-8.30 *	0.88
Family Gathering	56.72 ***	11.99 *	-7.43	3.41	36.54 ***
Social & Sports	-29.13 **	12.21	-4.54 *	66.77 ***	-14.62
Greetings & Celebrations	-11.58 ***	23.62 ***	-7.64	-28.64 ***	18.51
Friends & Family	-1.37	-5.98	-10.48	-12.79 ***	-9.33
Activities & Interests	-0.38	6.10	-21.42	-16.39	-30.58 **

and *relatives*. Among personal and social concerns, they increased language relating to *achievement*, *home*, and *religion*.

– *Cluster C<sub>1</sub>* individuals did not significantly change affective, cognitive, and perceptive expressions. Among function words, they decreased *second person pronouns* use and increased *conjunction* and *inclusive* use. They also significantly increased social words, such as words relating to *family*, *friends*, and *home*. This aligns with their topical changes post-enrollment. Therefore, they did not significantly change non-content word usage, but significantly changed content word usage; i.e., they did not change “how” they write, but changed “what” they write about.

– *Cluster C<sub>2</sub>* individuals significantly decreased language relating to a majority of affective and cognitive expressions, including *anger*, *anxiety*, *negative and positive affect*, *causation*, *cognitive mechanics*, *percept*, and *see*. They decreased using *first-person pronouns*. In other function words, they decreased *present tense*, *article*, *verbs*, *inclusive*, *preposition*, and *relative* use. Again, in personal and social concerns, they decreased the use of *friends* and *family*. These psycholinguistic changes indicate that *C<sub>2</sub>* individuals inhibit sharing personal and self-expressive content or prefer to share more about public-facing and less subjective content. This could be a sign of self-regulation.

– *Cluster C<sub>3</sub>* individuals significantly decreased using several affective, cognitive, and perceptive attributes. They decreased using *first person singular pronouns*, suggesting lowered self-attentional focus; however, the use of *third person pronouns* significantly increased. They also decreased using many function words, including *adverbs*, *verbs*, and *prepositions*. In contrast to *C<sub>2</sub>*, *C<sub>3</sub>* showed decreased *negative affect* and *swear* words and increased *positive affect* and *inclusive* keywords. We also find an increase in social words, such as *family*, *humans*, and *social*. These could be a manifestation of *C<sub>3</sub>* participants wanting to self-present in a more socially desirable or positive way. Similar to *C<sub>2</sub>* individuals, *C<sub>3</sub>* individuals decreased sharing *work*-related content.

– *Cluster C<sub>4</sub>* individuals increased multiple affective expressions, including *anger*, *negative affect*, and *swear*, whereas decreased *positive affect* use. Most cognitive and perceptive categories did not change, except for a significant decrease in *negation* and *feel*. These participants showed decreased *first person singular pronouns* usage but increased *past tense* usage. Most other function words and social words did not significantly change, except there was a significant reduction in the use of *adverbs*, *preposition*, *relative*, and *bio*.

## 5.2 RQ2: Validation of Observer Effect on Psychological Traits

Now, we aim to explain our observations through theories relating to individual differences. For each cluster, we examine the psychological traits and evaluate the behavioral and linguistic changes as observed in the social media use, presumably subject to observer effect. We contextualize and interpret the findings by drawing upon the literature in psychology and behavioral science [85, 145, 148]. Table 10 summarizes our observations.

**Cluster C<sub>0</sub> (routine-oriented)** individuals significantly decreased posting immediately after enrollment; however, their posting behaviors got closer to expected behaviors over time. This behavior change could be explained by their traits of high conscientiousness, which is known to be associated with self-monitoring [148]. The behavioral amendments over time is a form of *habituation* explained in behavioral science [39]. Linguistically, they decreased the use of first-person singular pronouns and increased the use of first-person plural pronouns and posting about public-facing events, which together could be considered to be reduced self-attentional focus and increased collective-identity-based language and increased posting about events attended as a part of a group [45].

**Cluster C<sub>1</sub> (emotionally stable and innovative)** individuals significantly decreased posting in the immediate two weeks after enrollment, but their posting behaviors became closer to the expected behaviors subsequently. Their social media language increased in **sociality** after enrollment [58]. As noted earlier, their use of content words increased, but their linguistic style remained similar. A possible explanation of their observed behaviors could be based on Middleton et al.’s observation that individuals with higher cognitive ability are less likely to show psychological reactance [109]. Again, the increased use of family-related keywords is known to be associated with lower self-monitoring [85]. They likely employ lower self-monitoring skills, are less bothered by the aspect of being “observed”, and are comfortable to continue sharing their social and personal life on social media.

**Cluster C<sub>2</sub> (withdrawn, disagreeable, and prone to stress and irritability)** individuals decreased posting on social topics like food and drinks, sports, and social events. This is also reflected in their lowered psycholinguistic use of personal and social words such as family and friends. However, they increased their posting activity. Their higher volume of post-enrollment posting behavior could be associated with higher self-monitoring skills as per prior work [84]. They also received greater engagement of likes and comments—plausibly a function of heightened information seeking on social media, which is known to be associated with higher neuroticism [146], as also for *C<sub>2</sub>* individuals.

**Table 9: Independent-sample  $t$ -tests in pre- and post- enrollment psycholinguistic (LIWC) use per cluster (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). Significant values are shaded in blue for positive changes, i.e., higher average occurrence in post-enrollment, and red for negative changes, i.e., lower average occurrence in post-enrollment period.**

LIWC	$t$ -test					LIWC	$t$ -test				
	C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>		C <sub>0</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
<b>Affect</b>						<b>Lexical Density and Awareness</b>					
Anger	2.01	-1.07	-2.02	-1.02	4.00 ***	Adverb	-3.00 **	-0.44	0.61	-3.36 ***	-2.46
Anxiety	1.41	0.03	-2.27	-1.957	1.93	Article	0.10	1.94	-3.60 ***	0.27	-1.34
Negative Affect	0.83	-0.93	-2.60 **	-2.83 **	2.09	Verb	-4.78 ***	0.47	-2.77 **	-5.53 ***	-1.36
Positive Affect	0.08	1.18	-4.49 ***	1.30 *	-2.06	Auxiliary Verb	-4.61 ***	0.40	0.10	-7.00 ***	-1.30
Sadness	1.427	-0.42	1.52	-1.46	-0.61	Conjun.	1.82	2.43	3.01 **	0.88	-0.15
Swear	1.134	0.60	-0.12	-3.06 **	7.53 ***	Exclusive	1.05	-1.56	0.80	-1.92	-0.33
<b>Cognition</b>						<b>Personal and Social Concerns</b>					
Causation	0.234	0.87	-2.69 **	-1.97	0.20	Inclusive	2.17	2.99 **	-3.47 ***	3.32 ***	-1.53
Certainty	4.08 ***	1.91	-2.12	-1.11	0.28	Negation	-1.38	-1.09	-0.90	-4.73 ***	-2.68 **
Cognitive Mechanics	1.32	0.86	-3.80 ***	-0.80	-0.93	Preposition	1.47	-1.01	-3.27 **	-3.11 **	-2.28
Inhibition	-1.13	-1.37	-3.53 ***	-0.02	0.60	Quantifier	-2.34	0.96	0.71	-0.06	0.50
Discrepanc.	-1.20	-1.61	1.08	-0.05	-0.55	Relative	-2.20	-1.16	-3.65 ***	-1.57	-2.98 **
Tentative.	0.43	-1.17	1.79	-1.83	1.23	<b>Achievement</b>					
Feel	-2.31	0.87	-1.66	-3.12 **	-2.51	Achievement	3.28 **	-0.91	-2.61 **	0.14	-1.08
Hear	5.48 ***	0.50	2.39	1.27	1.41	Bio	1.57	2.57	0.09	-0.20	-2.77 **
Insight	-1.23	-0.141	-0.32	-2.39	0.90	Body	-1.72	0.74	1.03	-1.34	-1.73
Percept	-0.07	0.35	-4.74 ***	-1.23	-1.50	Death	0.43	1.99	-0.93	-0.162	-0.45
See	-2.31	-0.80	-4.77 ***	-1.41	-0.80	Family	1.14	2.64 **	-2.06	3.66 ***	0.29
<b>Interpersonal Focus</b>						<b>Friends</b>					
1st Person Singular	-7.29 ***	-1.00	-5.78 ***	-2.35	-4.17 ***	Friends	-2.08	0.35 *	-1.46	-2.01	-1.17
1st Person Plural	2.25 *	0.47	-2.34	1.86	1.31	Health	0.52	0.47	-0.34	-0.11	-0.81
2nd Person	-1.43	-3.32 ***	5.71 ***	1.16	-0.70	Home	3.14 **	2.77 **	-2.19	1.39	0.08
3rd Person	-0.12	-0.63	-0.26	4.61 ***	-0.03	Humans	-2.94 **	-1.67	0.51	2.85 **	-0.62
Indef. Pronoun	-3.29 **	-1.33	-1.30	-3.43 ***	0.92	Money	-1.81	-1.06	-1.05	1.01	-1.24
Future Tense	0.32	-0.65	2.32	-0.69	-0.36	Religion	2.29	2.48	-1.06	-0.87	-0.14
Past Tense	1.85	0.28	-1.99	-0.158	2.61 **	Sexual	1.62	-0.27	1.07	-0.62	0.56
Present Tense	-5.54 ***	0.19	-3.15 **	-6.49 ***	-1.90	Social	-1.02	-0.63	-1.54	2.79 **	0.30
						Work	0.29	-0.58	-4.57 ***	-2.96 **	-1.74

**Cluster C<sub>3</sub> (positive, friendly, and well-balanced)** individuals increased posting after enrollment. Extraversion is known to positively correlate with public self-consciousness [155] and self-monitoring [17]. Similar to C<sub>2</sub>, greater posting behavior in C<sub>3</sub> could be manifested by high self-monitoring skills [84]. Further, high conscientiousness could also indicate a desire to appear as “good” participants or self-present in a more desirable way [16]—this could be reflected in their increased social media activities, increased positive affect, and decreased negative affect and swear words, as explained by the self-presentation literature [71, 86]. High agreeableness is known to be associated with people’s likelihood to seek acceptance and maintain social connections [146]. A similar phenomenon is observable for them as their posts elicited a greater number of likes and comments compared to before enrollment.

**Cluster C<sub>4</sub> (curious and adventurous)** individuals did not significantly change posting behaviors immediately but significantly increased posting over time. They also showed significant linguistic changes in the post-enrollment period. They increased posting about many social aspects of life despite significantly reducing the use of first-person singular pronouns and many function words. They lowered the use of negations and exclusives, suggesting lowered cognitive complexity in language—which could be

associated with less personal content [117]. These changes suggest that C<sub>4</sub> individuals are likely to self-regulate their social media use to present selective aspects of life without sharing too intimate content. Also, greater openness is associated with high psychological reactance [145], which could manifest in detached sharing about personal and first-person singular content. Openness is associated with greater resiliency and externally induced behavioral changes [110]; however, its interplay with observer effect remains to be examined further.

## 6 ROBUSTNESS OF FINDINGS

### 6.1 Placebo Tests

Our study design considers the enrollment in the study as *treatment*. We need to ensure that the observed effects are actually *caused* by the treatment, and not due to other confounds or a chance. So, we conduct placebo tests [147] drawing on permutation test approaches from prior work [9, 37, 133]. We permute (randomize) several *placebo* dates within the pre-enrollment data. Here, the placebo tests are meant to rule out the likelihood that significant changes in social media use could also happen around dates other than the enrollment date (or placebo dates). We assign 150 placebo dates, and

**Table 10: Summary for each cluster, their traits, and observed changes in behaviors, topics, and psycholinguistics.**

	Traits	Behavior	Topics	Psycholinguistics	Notes / Descriptor
C <sub>0</sub>	High (Conscientiousness, Sleep Quality), Low (Openness, Cognitive Ability)	Posting significantly reduces in the initial few days, then back to expected behaviors (Figure 7a)	Increased sharing about public-facing information (Table 8)	Increased (anger, achievement, home, religion), Decreased (feel, first person singular, present tense, function words, friends, humans) (Table 9)	High conscientiousness is associated with self-monitoring. Habituation in posting behavior. Decreased self-attentional focus.
C <sub>1</sub>	High (Cognitive Ability), Low (Neuroticism)	Posting significantly decreased in the first two weeks, then closer to expected behaviors (Figure 7c)	Increased sharing about family gathering, social, and online greeting related activities (Table 8)	Increased (social words), Decreased (2nd person) (Table 9)	These participants are trait-wise more reasonable and composed. They show high sociality post-enrollment. They show low psychological reactance and low self-monitoring skills, and are less bothered about being “observed”.
C <sub>2</sub>	Low (Extraversion, Agreeableness, Conscientiousness, PA, Sleep Quality), High (Neuroticism, Cognitive Ability, NA, Anxiety)	Posting significantly increased throughout. Received greater engagement. (Figure 7e, 7f)	Decreased sharing about food and social topics (Table 8)	Increased (hear, future tense), Decreased (affective, cognitive, perceptive, 1st person pronouns, function words, social words) (Table 9)	Trait-wise, they may be more withdrawn, and prone to stress and irritability. High self-monitoring skills, and heightened information seeking (associated with high neuroticism).
C <sub>3</sub>	High (Extraversion, Agreeableness, Conscientiousness, PA, Sleep Quality), Low (Neuroticism, NA, Anxiety)	Posting activity significantly increases throughout. Received greater engagement. (Figure 7g, 7h)	Decreased sharing about personal events (Table 8)	Increased (social words, third person pronouns), Decreased (affective, cognitive, perceptive, first-person pronouns, function words) (Table 9)	They intend to self-present in a more desirable way. Likelihood to seek acceptance and maintain social connections.
C <sub>4</sub>	High Openness	No immediate significant difference in posting frequency, but posting significantly increases throughout. More likes received. (Figure 7i, 7j)	Decreased sharing about news and holiday plans. Increased sharing about food/family gathering (Table 8)	Increased (anger, NA, swear, past tense), Decreased (PA, negation, feel, 1st person singular, function words) (Table 9)	They show self-regulation. Share lesser personal-content. High psychological reactance manifested in detached sharing about personal content.

repeat the same time series comparison around the placebo dates—for every placebo date, we compute the  $t$ -tests in the post-placebo date actual and predicted time series data. Then, over all the permutations of placebo dates, we compute the probability ( $p$ -value) of significant differences around placebo dates. A  $p$ -value lower than 0.05 would reject the null hypothesis that the significance is by chance, also revealing the credibility of any significant changes observed around the (real) enrollment date.

Out of 150 permutations, C<sub>0</sub> and C<sub>4</sub> show significance in 2 and 1 permutations, respectively, and the other three clusters show no significant permutations. Therefore, the probability of a significant placebo effect is close to 0 for all the clusters, revealing that the significance observed around the *actual enrollment dates* (or *treatment*) is not by chance. This test also validates our extrapolation of expected behaviors in the post-enrollment period.

## 6.2 Sensitivity Analysis on Data Availability Timeline

We also conduct a sensitivity analysis by varying the threshold of the availability of minimum post-enrollment data (15 days, 30 days, 45 days) to see if the quantity of available data introduced any biases in our findings. For each of the other thresholds, we repeat the experiments for 365, 344, and 335 participants, respectively; however, the findings do not significantly change compared to what we have for the 60-day threshold. In addition, Cox Proportional-Hazards regression models [26, 47] for all the examined measures (posts made and engagements received) confirm no statistical significance with respect to the quantity of time of data in the post-enrollment period. This suggests that our findings are not sensitive to the minimum threshold of post-enrollment data considered.

## 7 DISCUSSION

This work provides insights into how the observer effect occurs, how long it lasts, and how its occurrences vary across individuals. Theoretically, this work advances our knowledge about how participants varying in psychological traits could change social media use differently in prospective research design settings. These behavioral changes are explained by behavioral science and psychology theories, including self-monitoring [152], public self-consciousness [11], and psychological reactance [28]. Methodologically, this work contributes a computational and causal framework for modeling observer effect in prospective research studies in general, and those involving the monitoring of social media use in particular. Our work is motivated by person-centered approaches of clustering individuals on psychological traits and studying the behavior changes per cluster [168]. A strength of person-centered approach is that it views each cluster as an integrated totality [67, 168], and helps us draw within-person (or within-clusters, here) insights and interpretations, i.e., given an individual with a certain combination of traits, how they would likely behave after an intervention. We discuss this work’s implications in recommending strategies to correct for biases arising from the observer effect in social media studies.

### 7.1 Theoretical Implications

**7.1.1 Observer effect and behavior change research.** This study advances our knowledge in observer effect research. Typically, the observer effect has been hard to study because researchers could only access data generated after participant recruitment [106]. This has precluded researchers from measuring observer effect since it necessitates access to and comparison with a subject’s otherwise normative and non-observed behavior (e.g., prior to enrollment in

the prospective study), or the counterfactual how they would have behaved without the presence of an observer. In addition, there is no established gold standard for measuring observer effect. To the best of our knowledge, this is the first study measuring the observer effect of social media use. The longitudinal and historical nature of the social media data stream allowed access to extended periods of an individual's behavior on the platform, including pre-enrollment data. This enabled us to build behavioral models on typical or expected behaviors, which we leveraged in this work.

We also note an interesting aspect. While the lack of a control group in our study can be seen as a limitation, it, in fact, supports our design choice to overcome a known challenge with studying the observer effect. Historically, the observer effect has been challenging to study in research because of a paradox that if any control/comparison group is enrolled, they also inherently get subjected to the observer effect [115], or the John Henry effect [5, 139]. Adopting a synthetic control-based approach enabled us to overcome this limitation and examine the observer effect within the enrolled participants. Further, placebo tests help comparison within different timeframes (before the study) and mitigate temporal confounds, ensuring statistical significance and rigor of our examinations.

Social media experiments are also unique in comparison to traditional experiments. For instance, social media experiments are sensitive to people's conscious choices and decisions about using these platforms. That is to say, social media use happens in a naturalistic setting, an intentional and conscious behavior that individuals can alter at their will. The likelihood of behavior change attributed to observer effect increases for conscious behaviors [27], as explained in prior research—Arkin and Shepperd noted self-consciousness influences one's strategic self-presentation and Snyder noted people are likely to self-monitor their self-presentations, expressive behaviors, and non-verbal affective displays [11, 152]. These are relevant and important aspects of social media use. Additionally, social media use comprises "social activity" and verbal and expressive behaviors. In contrast, traditional experiments primarily comprise personal activities undertaken in somewhat non-natural or even artificial settings. These differences together warrant studying the observer effect in social media experiments.

**7.1.2 Correcting biases in prospective use of social media as a passive sensor.** This study provides insights regarding the prevalence and degree of the observer effect in social media use by psychological traits of participants. We draw a novel understanding of how people with different combinations of these traits could behave when subjected to the observer effect. These findings inform research about correcting data, biases, and models when implementing practical and prospective data-driven assessments and interventions. In this regard, this study contributes to the recommendations by Ruths and Pfeffer in correcting biases of big data technologies [130]. Specifically, this study helps us to gauge what to expect when social media data is used to assess human behaviors in a prospective setting. For instance, this work informs us that composed and reasonable individuals (Cluster  $C_1$ ) are likely to decrease posting in the immediate period but might show habituation or return to expected behaviors over time, whereas those with high openness (Cluster  $C_4$ ) may not show any immediate change but increase posting over a period of time. These findings help us be more cognizant about

which individuals might significantly deviate from their otherwise expected behaviors and accordingly build personalized models that are robust to people's baseline traits and tendencies to be impacted by the observer effect.

**7.1.3 Generating testable hypotheses.** Our findings can also help to generate hypotheses relating to the observer effect in social media. For example, a reduction in the use of first-person pronouns signals a presence of the observer effect. Additionally, in Section 5.2, we explain the findings through theories in psychology and behavioral science literature. These associations can be formulated as testable hypotheses in future research. For instance, how self-regulation and self-monitoring associates with observer effect. Future research can incorporate other intrinsic and social processes, such as self-censorship and privacy perceptions, which may also interact with social media behavioral change [49, 104].

Due to the lack of direct means to measure the success and construct validity of this work, we evaluated and situated the findings by referring to existing theories. While our work targeted to obtain passive and objective forms of assessment, it is also interesting to examine self-reported assessments about the observer effect. Therefore, this work motivates us to design and conduct surveys and interviews, which would help us gauge complementary information about how the observer effect manifests in social media behavior.

**7.1.4 Self-selection and "who is the observer?"** In this study, observers were a group of researchers with whom the participants willingly consented to their data based on a data-sharing protocol. These participants self-selected themselves in the study, for which they were compensated. Our participant pool was U.S. information workers. While this enabled us to study on a population with comparable familiarity with computing technologies, the observer effect can occur more broadly. For example, the observer effect may differentially manifest in populations with varying digital and privacy literacy as well as socio-cultural attitudes towards these issues [44, 142]. While our findings may not necessarily generalize to other populations, our study design and computational approach can be repurposed. The observer effect can also occur in other scenarios involving a variety of observers and data-sharing terms, such as clinicians observing patients' health, employers observing workers' productivity, or social media platforms monitoring user activities to control policy violations. Olteanu et al. [114] connected "online" observer effect with people's disclosure behaviors in terms of how individuals are more likely to share unpopular, sensitive, and more personal opinions in private and anonymous spaces than public ones [21, 141, 149]. In addition, the observer effect may occur differently in the cases of anonymous or pseudonymous settings. While anonymity might help an individual for greater intimate self-disclosure [52, 116], it remains to be empirically examined how the observer effect can interact with anonymity and have a trickle-down effect on people's self-disclosure on social media. Therefore, it is important to understand what factors influence the observer effect in the real world.

## 7.2 Implications for Researchers & Practitioners

This research showed that individuals who deviated from their expected behaviors when subjected to real-time and prospective data

collection settings — attributed as some form of observer effect. This effect needs to be accounted for to successfully instrument real-time applications using social media to derive behavioral or psychological assessments. The computational framework adopted in this study can be used to measure observer effects in various contexts. Researchers can use such approaches to identify cases of observer effect-based deviations and build predictive models robust to such effects in a person-centric fashion. This study reveals that self-reported psychological traits can not only be used to stratify and cluster individuals, but also to explain behavioral changes due to the observer effect. Similar approaches can be used to build person-centric models of correction for different groups of individuals. Relatedly, we noted in the Introduction how most social media-based studies of human behaviors are retrospective and observational. However, a major implication of this research body is to inform practical and real-time interventions. Our work implies that it is worth revisiting the retrospective analyses along with corrections for the observer effect before significant efforts and resources are invested in making the interventions.

Besides highlighting the potential methodological biases, this study also reinforces an ethical question about social media research of human behaviors in general (both retrospective and prospective). It motivates us to critically reflect and rethink the implications surrounding individuals' autonomy and comfort in using social media platforms. People primarily use social media to share and connect with others. However, if external interventions interfere with their social media use or make them feel uncomfortable or surveilled—as revealed to be the same for at least some participants in this study—then the fundamental goals and expectations of using social media platforms can be compromised. Such an unintended consequence needs to be evaluated by researchers, practitioners, as well as the owners of social media platforms. To this end, this work encourages us to critique the trade-offs between the harms and benefits of using social media-based technologies for deriving psychological assessments, and also reinforces the necessity of consenting to individuals' social media data and their specific use.

### 7.3 Limitations and Future Directions

It is also important to note how our findings are an artifact of the domain and the participant pool. This study is conducted on a specific participant pool of information workers in the context of workplace settings. Such a factor may affect the changes observed in the work-related language (in Table 8 and Table 9). In addition, our study is not devoid of biases due to self-selection [114], and our work adopts a person-centered approach to somewhat mitigate this challenge [89]. While our clustering-based approach helped us examine and understand how observer effect impacts different individuals' social media use, our study population does not include all possible combinations of psychological traits. Future experiments can explore more conclusive and generalizable evidence about the observer effect and whether these are opportunities or challenges in other situations and contexts. We also note that even though it would have been interesting (and possibly more accurate) to include the demographic attributes of individuals in clustering, we excluded these attributes to primarily steer

away from “demographic profiling” related interpretations and ethical concerns—demographic attribute-based stratified modeling has been associated with reinforcing and exacerbating stereotypes and existing societal biases [88, 119]. In addition, given that our dataset is not representative of all demographic and marginalized groups, the non-demographic psychological traits are more robust for studying as well as for reproducibility and applicability of research.

Also, our study did not include a comparison with a control group in measuring the observer effect. The lack of a control group can be considered both a limitation as well as strength of the study design. The phenomenon of observer effect has been generally challenging to study because of a paradox that any enrolled control group is also inherently subjected to the observer effect [115], or the John Henry Effect [5, 139]. Simultaneously, studying participants' data without consent would raise ethical questions [24, 90]. Future research can adopt alternative study designs that include between-individual analyses. Such an approach can recruit a control group at a later date after recruiting the experimental group or have staggered recruitments of participants and then compare the experimental group's with the control group's pre-enrollment data.

It is important to note that social media use significantly evolves over time both within and across platforms [10]. Our study employed Facebook data until 2019 (pre-pandemic), and future work can investigate how these observations may change over the years and on different platforms. For instance, several social media platforms have transformed over the years in terms of both user base and usage [12]. In addition, the physical isolation since the pandemic might have influenced people's social media use compared to before. Such an external factor can confound examinations of the observer effect, and thereby, needs to be accounted for. Additionally, the pandemic, as well as the emerging trend of Generative AI technologies (e.g., ChatGPT, conversational search, etc.) have likely influenced people's interactions with AI and social technologies. Content has also been evolving; for example, ephemeral content (stories on Facebook, Instagram, Whatsapp, etc.) and video-based content (Instagram reels, YouTube shorts, TikTok Videos, etc.) have become increasingly prevalent, and platforms also provide audience control-like features [61]. Accordingly, studying the variation of audience control would be an important dimension to consider in future work in the specific context of the observer effect.

## 8 CONCLUSION

We examined the likelihood and degree of the observer effect in longitudinal social media use. We operationalized the observer effect in two dimensions of social media (Facebook) use—behavioral and linguistic changes. Participants consented to Facebook data collection over an average retrospective period of 82 months and an average prospective period of 5 months around the enrollment date of our study. We adopted a synthetic control-based causal approach to measure how people deviated from expected social media use after enrollment. We obtained expected use by extrapolating from historical use using time-series (ARIMA) forecasting. We found that the deviation in social media use varies across individuals based on psychological traits. Individuals with high cognitive ability and low neuroticism immediately decreased posting after enrollment, and



those with high openness significantly increased posting. Linguistically, most individuals decreased the use of first-person pronouns, reflecting lowered sharing of intimate and self-attentional content. While some increased posting about public-facing events, others increased posting about family and social gatherings. We validated our observations based on psychological traits drawing from psychology and behavioral science theories, such as self-monitoring, public self-consciousness, and self-presentation. The findings provide recommendations to correct observer effects in social media data-driven assessments of human behavior.

## 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 Nick Jaffe, Jordyn Seybolt, Chris Martin, and the members of the Tesseract team and SocWeb lab for contributing to and providing feedback on this work.

## REFERENCES

- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association* 105, 490 (2010), 493–505.
- Alberto Abadie and Javier Gardeazabal. 2003. The economic costs of conflict: A case study of the Basque Country. *American economic review* 93, 1 (2003).
- Mohammad-Ali Abbasi, Sun-Ki Chai, Huan Liu, and Kiran Sagoo. 2012. Real-world behavior analysis through a social media lens. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer.
- Alessandro Acquisti and Ralph Gross. 2006. Imagined communities: Awareness, information sharing, and privacy on the Facebook. In *International workshop on privacy enhancing technologies*. Springer, 36–58.
- John G Adair. 1984. The Hawthorne effect: a reconsideration of the methodological artifact. *Journal of applied psychology* 69, 2 (1984), 334.
- Icek Ajzen. 1985. From intentions to actions: A theory of planned behavior. In *Action control*. Springer, 11–39.
- Icek Ajzen et al. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2 (1991), 179–211.
- C Norman Alexander Jr and Pat Lauderdale. 1977. Situated identities and social influence. *Sociometry* (1977), 225–233.
- 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.
- A Archambault and J Grudin. 2012. A Longitudinal Study of Facebook, LinkedIn, and Twitter Use. In *Proceedings of CHI 2012*.
- Robert M Arkin and James A Shepperd. 1990. Strategic self-presentation: An overview. (1990).
- Brooke Auxier and Monica Anderson. 2021. Social media use in 2021. *Pew Research Center* 1 (2021), 1–4.
- Albert Bandura. 1988. Organisational applications of social cognitive theory. *Australian Journal of management* 13, 2 (1988), 275–302.
- Albert Bandura. 2002. Social cognitive theory in cultural context. *Applied psychology* 51, 2 (2002), 269–290.
- Susan B Barnes. 2006. A privacy paradox: Social networking in the United States. *First Monday* 11, 9 (2006).
- 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.
- Murray R Barrick, Laura Parks, and Michael K Mount. 2005. Self-monitoring as a moderator of the relationships between personality traits and performance. *Personnel psychology* 58, 3 (2005), 745–767.
- Sebastian Bauhoff. 2014. The effect of school district nutrition policies on dietary intake and overweight: a synthetic control approach. *Economics & Human Biology* 12 (2014), 45–55.
- Anne Beaulieu. 2010. Research Note: From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge. *Social Studies of Science* 40, 3 (2010), 453–470.
- Marshall H Becker. 1974. The health belief model and sick role behavior. *Health education monographs* 2, 4 (1974), 409–419.
- Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 2011. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In *International Conference on Weblogs and Social Media (ICWSM)*.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *JMLR* 3, Jan (2003), 993–1022.
- Shelley Boulianne. 2015. Social media use and participation: A meta-analysis of current research. *Information, communication & society* (2015).
- Danah Boyd. 2016. Untangling research and practice: What Facebook's "emotional contagion" study teaches us. *Research Ethics* 12, 1 (2016), 4–13.
- 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.
- Mike J Bradburn, Taane G Clark, Sharon B Love, and Douglas Graham Altman. 2003. Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *British journal of cancer* 89, 3 (2003), 431–436.
- Augustine Brannigan and William Zwermer. 2001. The real "Hawthorne effect".
- Jack W Brehm. 1966. A theory of psychological reactance. (1966).
- Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L Scott. 2015. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics* (2015), 247–274.
- Kay H Brodersen, Alain Hauser, and Maintainer Alain Hauser. 2017. Package CausalImpact. *Google LLC: Mountain View, CA, USA* (2017).
- Maira 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*.
- Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian* 17 (2018).
- Ralph Catalano. 1979. *Health, behavior and the community: An ecological perspective*. Pergamon Press New York.
- Junghoon Chae, Dennis Thom, Yun Jang, SungYe Kim, Thomas Ertl, and David S Ebert. 2014. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics* 38 (2014), 51–60.
- Stevie Chancellor and Mummun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 1–11.
- Stevie Chancellor, Zhiyuan (Jerry) Lin, Erica Goodman, Stephanie Zerwas, and Mummun De Choudhury. 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*. ACM. in press.
- Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction* CSCW (2017).
- Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Neural information processing systems*, Vol. 22. Citeseer, 288–296.
- Luke F Chen, Charlene Carriker, Russell Staheli, Pamela Isaacs, Brandon Elliott, Becky A Miller, Deverick J Anderson, Rebekah W Moehring, Sheila Vereen, Judie Bringhurst, et al. 2013. Observing and improving hand hygiene compliance implementation and refinement of an electronic-assisted direct-observer hand hygiene audit program. *Infection Control & Hospital Epidemiology* 34, 2 (2013).
- Luke F Chen, Mark W Vander Weg, David A Hofmann, and Heather Schacht Reisinger. 2015. The Hawthorne effect in infection prevention and epidemiology. *Infection control & hospital epidemiology* 36, 12 (2015), 1444–1450.
- Kendra Cherry and Paul G Mattiuzzi. 2010. *The Everything Psychology Book: Explore the human psyche and understand why we do the things we do*. Simon and Schuster.
- Mecca Chiesa and Sandy Hobbs. 2008. Making sense of social research: How useful is the Hawthorne Effect? *European Journal of Social Psychology* 38, 1 (2008), 67–74.
- Prerna Chikersal, Danielle Belgrave, Gavin Doherty, Angel Enrique, Jorge E Palacios, Derek Richards, and Anja Thieme. 2020. Understanding Client Support Strategies to Improve Clinical Outcomes in an Online Mental Health Intervention. In *Proc. CHI*.
- SoeYoon Choi. 2023. Privacy literacy on social media: Its predictors and outcomes. *International Journal of Human-Computer Interaction* 39, 1 (2023).
- 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.

- [46] Paul T Costa Jr, Jeffrey H Herbst, Robert R McCrae, Jack Samuels, and Daniel J Ozer. 2002. The replicability and utility of three personality types. *European Journal of Personality* 16, S1 (2002), S73–S87.
- [47] David R Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 2 (1972), 187–202.
- [48] Aron Culotta. 2014. Estimating county health statistics with Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1335–1344.
- [49] Sauvik Das and Adam Kramer. 2013. Self-censorship on Facebook. In *Proc. ICWSM*.
- [50] Vedant Das Swain, Koustuv Saha, Hemang Rajvanshy, Anusha Sirigiri, Julie M Gregg, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *Proc. IMWUT* (2019).
- [51] 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.
- [52] 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*.
- [53] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *ICWSM*.
- [54] 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.
- [55] 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.
- [56] Judith S Donath. 1999. Identity and deception in the virtual community. *Communities in cyberspace* 1996 (1999), 29–59.
- [57] Brooke Erin Duffy and Ngai Keung Chan. 2019. “You never really know who’s looking”: Imagined surveillance across social media platforms. *New Media & Society* 21, 1 (2019), 119–138.
- [58] Sindhu Kiranmai Ernala, Kathan H Kashiparekh, Amir Bolous, Ali Asra, John M Kane, Michael L Birnbaum, and Munmun De Choudhury. 2021. A Social Media Study on Mental Health Status Transitions Surrounding Psychiatric Hospitalizations. *PACM HCI CSCW* (2021).
- [59] 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*.
- [60] 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.
- [61] Sindhu Kiranmai Ernala, Stephanie S Yang, Yuxi Wu, Rachel Chen, Kristen Wells, and Sauvik Das. 2021. Exploring the Utility Versus Intrusiveness of Dynamic Audience Selection on Facebook. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–30.
- [62] Ruth Faden, Nancy Kass, Danielle Whicher, Walter Stewart, and Sean Tunis. 2013. Ethics and informed consent for comparative effectiveness research with prospective electronic clinical data. *Medical Care* (2013), S53–S57.
- [63] Casey Fiesler and Nicholas Proferes. 2018. “Participant” perceptions of Twitter research ethics. *Social Media+ Society* (2018).
- [64] Martin Fishbein and Icek Ajzen. 1980. Understanding attitudes and predicting social behavior. (1980).
- [65] Martin Fishbein and Joseph N Cappella. 2006. The role of theory in developing effective health communications. *Journal of communication* 56 (2006), S1–S17.
- [66] William A Fisher, Jeffrey D Fisher, and Jennifer Harman. 2003. The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior. *Social psychological foundations of health and illness* 22 (2003), 82–106.
- [67] Roseanne J Foti, Nicole J Thompson, and Sarah F Allgood. 2011. The pattern-oriented approach: A framework for the experience of work. *Industrial and Organizational Psychology* 4, 1 (2011), 122–125.
- [68] Nathan S Fox, Jennifer S Brennan, and Stephen T Chasen. 2008. Clinical estimation of fetal weight and the Hawthorne effect. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 141, 2 (2008), 111–114.
- [69] Chad Fulton. 2015. Estimating time series models by state space methods in Python: Statsmodels.
- [70] Saby Ghoshray. 2013. Employer surveillance versus employee privacy: The new reality of social media and workplace privacy. *N. Ky. L. Rev.* 40 (2013), 593.
- [71] Erving Goffman. 1959. The presentation of self in everyday life. (1959).
- [72] 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).
- [73] Crystal Gong, Koustuv Saha, and Stevie Chancellor. 2021. “The Smartest Decision for My Future”: Social Media Reveals Challenges and Stress During Post-College Life Transition. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.
- [74] Emma Grace, Parimala Raghavendra, Julie M McMillan, and Jessica Shipman Gunion. 2019. Exploring participation experiences of youth who use AAC in social media settings: Impact of an e-mentoring intervention. *Augmentative and Alternative Communication* 35, 2 (2019), 132–141.
- [75] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Demographics of Social Media Users in 2016. [pewinternet.org/2016/11/11/social-media-update-2016/](http://pewinternet.org/2016/11/11/social-media-update-2016/). Accessed: 2017-02-12.
- [76] Pamela Grimm. 2010. Social desirability bias. *Wiley international encyclopedia of marketing* (2010).
- [77] Bernard Guerlin. 2010. Social facilitation. *The Corsini encyclopedia of psychology* (2010), 1–2.
- [78] Maria-Dolores Guillamón, Ana-María Ríos, Benedetta Gesuele, and Concetta Metallo. 2016. Factors influencing social media use in local governments: The case of Italy and Spain. *Government Information Quarterly* 33, 3 (2016), 460–471.
- [79] Jamie Guillory and Jeffrey T Hancock. 2012. The effect of LinkedIn on deception in resumes. *Cyberpsychology, Behavior, and Social Networking* (2012).
- [80] 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*.
- [81] Sharath Chandra Guntuku, J Russell Ramsay, Raina M Merchant, and Lyle H Ungar. 2019. Language of ADHD in adults on social media. *Journal of attention disorders* (2019).
- [82] Sharath Chandra Guntuku, H Andrew Schwartz, Adarsh Kashyap, Jessica S Gaulton, Daniel C Stokes, David A Asch, Lyle H Ungar, and Raina M Merchant. 2020. Variability in Language used on Social Media prior to Hospital Visits. *Scientific Reports* 10, 1 (2020), 1–9.
- [83] Oliver L Haimson, Albert J Carter, Shanley Corvite, Brookelyn Wheeler, Lingbo Wang, Tianxiao Liu, and Alexxus Lige. 2021. The major life events taxonomy: Social readjustment, social media information sharing, and online network separation during times of life transition. *Journal of the Association for Information Science and Technology* (2021).
- [84] Jeffrey A Hall and Natalie Pennington. 2013. Self-monitoring, honesty, and cue use on Facebook: The relationship with user extraversion and conscientiousness. *Computers in Human Behavior* 29, 4 (2013), 1556–1564.
- [85] Qiwei He, Cees AW Glas, Michal Kosinski, David J Stillwell, and Bernard P Veldkamp. 2014. Predicting self-monitoring skills using textual posts on Facebook. *Computers in Human Behavior* 33 (2014), 69–78.
- [86] 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.
- [87] John D Holden. 2001. Hawthorne effects and research into professional practice. *Journal of evaluation in clinical practice* 7, 1 (2001), 65–70.
- [88] 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.
- [89] Katy Jordan. 2018. Validity, reliability, and the case for participant-centered research: Reflections on a multi-platform social media study. *International Journal of Human-Computer Interaction* 34, 10 (2018), 913–921.
- [90] 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).
- [91] 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.
- [92] Anna Kawakami, Shreya Chowdhary, Shamsi T Iqbal, Q Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. 2023. Sensing Wellbeing in the Workplace, Why and For Whom? Envisioning Impacts with Organizational Stakeholders. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* (2023).
- [93] Alan E Kazdin. 1982. Observer effects: Reactivity of direct observation. *New Directions for Methodology of Social & Behavioral Science* (1982).
- [94] Alan E Kazdin. 2011. Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist* 66, 8 (2011), 685.
- [95] Emre Kiciman, Scott Counts, and Melissa Gasser. 2018. Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use. In *ICWSM*. 171–180.
- [96] Holly Korda and Zena Itani. 2013. Harnessing social media for health promotion and behavior change. *Health promotion practice* 14, 1 (2013).
- [97] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. (2013).
- [98] 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.

- [99] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The parable of Google Flu: traps in big data analysis. *Science* 343, 6176 (2014).
- [100] Margaret D LeCompte and Judith Preissle Goetz. 1982. Problems of reliability and validity in ethnographic research. *Review of educational research* (1982).
- [101] Jason Liu, Elissa R Weitzman, and Rumi Chunara. 2017. Assessing behavioral stages from social media data. In *CSCW*.
- [102] 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.
- [103] 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.
- [104] 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.
- [105] Stephen M Mattingly, Julie M Gregg, Pino Audia, Ayse Elvan Bayraktaroglu, Andrew T Campbell, Nitesh V Chawla, et al. 2019. The Tesseract Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. (2019).
- [106] Jim McCambridge, John Witton, and Diana R Elbourne. 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.
- [107] David McDowall, Richard McCleary, and Bradley J Bartos. 2019. *Interrupted time series analysis*. Oxford University Press.
- [108] Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 889–892.
- [109] Jay Middleton, Walter Buboltz, and Barlow Soper. 2015. The relationship between psychological reactance and emotional intelligence. *The Social Science Journal* 52, 4 (2015), 542–549.
- [110] Claude H Miller, Michael Burgooon, Joseph R Grandpre, and Eusebio M Alvaro. 2006. Identifying principal risk factors for the initiation of adolescent smoking behaviors: The significance of psychological reactance. *Health communication* 19, 3 (2006), 241–252.
- [111] Hugh Miller. 1995. The presentation of self in electronic life: Goffman on the Internet. In *Embodied knowledge and virtual space conference*, Vol. 9.
- [112] Torin Monahan and Jill A Fisher. 2010. Benefits of ‘observer effects’: lessons from the field. *Qualitative research* 10, 3 (2010), 357–376.
- [113] Ryan Olson, Jessica Verley, Lindsey Santos, and Coresta Salas. 2004. What we teach students about the Hawthorne studies: A review of content within a sample of introductory IO and OB textbooks. *The Industrial-Organizational Psychologist* 41, 3 (2004), 23–39.
- [114] 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.
- [115] David Oswald, Fred Sherratt, and Simon Smith. 2014. Handling the Hawthorne effect: The challenges surrounding a participant observer. *Review of social studies* 1, 1 (2014), 53–73.
- [116] Sai Teja Peddinti, Keith W Ross, and Justin Cappos. 2014. “On the internet, nobody knows you’re a dog” a twitter case study of anonymity in social networks. In *Proceedings of the second ACM conference on Online social networks*. 83–94.
- [117] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.
- [118] James O Prochaska, Sara Johnson, and Patricia Lee. 1998. The transtheoretical model of behavior change. (1998).
- [119] 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.
- [120] Liana Razmerita, Kathrin Kirchner, and Pia Nielsen. 2016. What factors influence knowledge sharing in organizations? A social dilemma perspective of social media communication. *Journal of knowledge Management* (2016).
- [121] 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.
- [122] Press Release. 2016. Census Bureau Reports. <https://www.census.gov/newsroom/press-releases/2016/cb16-139.html>. Accessed: 2020-04-04.
- [123] Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. 99–107.
- [124] 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.
- [125] Jon E Roecoeklein. 1998. *Dictionary of theories, laws, and concepts in psychology*. Greenwood Publishing Group.
- [126] Fritz Jules Roethlisberger and William J Dickson. 2003. *Management and the Worker*. Vol. 5. Psychology press.
- [127] Irwin M Rosenstock. 1974. Historical origins of the health belief model. *Health education monographs* 2, 4 (1974), 328–335.
- [128] Irwin M Rosenstock, Victor J Strecher, and Marshall H Becker. 1988. Social learning theory and the health belief model. *Health education quarterly* 15, 2 (1988), 175–183.
- [129] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.
- [130] Derek Ruths and Jürgen Pfeffer. 2014. Social media for large studies of behavior. *Science* 346, 6213 (2014), 1063–1064.
- [131] Koustuv Saha, Ayse E Bayraktaroglu, Andrew T Campbell, Nitesh V Chawla, Munmun De Choudhury, Sidney K D’Mello, Anind K Dey, Ge Gao, Julie M Gregg, Krithika Jagannath, et al. 2019. Social media as a passive sensor in longitudinal studies of human behavior and wellbeing. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [132] Koustuv Saha, Ted Grover, Stephen M 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–32.
- [133] Koustuv Saha, Manikanta D Reddy, Vedant das Swain, Julie M Gregg, Ted Grover, Suwen Lin, Gonzalo J Martinez, Stephen M Mattingly, Shayan Mirjafari, Raghu Mulukutla, et al. 2019. Imputing missing social media data stream in multisensor studies of human behavior. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 178–184.
- [134] Koustuv Saha, Jordyn Seybolt, Stephen M Mattingly, Talayah 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 *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–22.
- [135] Koustuv Saha and Amit Sharma. 2020. Causal Factors of Effective Psychosocial Outcomes in Online Mental Health Communities. In *ICWSM*.
- [136] Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kiciman, and Munmun De Choudhury. 2019. A Social Media Study on the Effects of Psychiatric Medication Use. In *ICWSM*.
- [137] 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*.
- [138] Koustuv Saha, Asra Yousuf, Ryan L Boyd, James W Pennebaker, and Munmun De Choudhury. 2022. Social media discussions predict mental health consultations on college campuses. *Scientific reports* 12, 1 (2022), 123.
- [139] Gary Saretsky. 1972. The OEO PC experiment and the John Henry effect. *The Phi Delta Kappan* 53, 9 (1972), 579–581.
- [140] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. Finding a “kneede” in a haystack: Detecting knee points in system behavior. In *ICDCS*.
- [141] Sarita Yardi Schoenebeck. 2013. The secret life of online moms: Anonymity and disinhibition on youbemo.com. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- [142] Lara Schreurs and Laura Vandenbosch. 2021. Introducing the Social Media Literacy (SMILE) model with the case of the positivity bias on social media. *Journal of Children and Media* 15, 3 (2021), 320–337.
- [143] 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.
- [144] Skipper Seabold and Josef Perktold. 2010. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, Vol. 57. Austin, TX, 10–25080.
- [145] Eric A Seemann, Walter C Buboltz, Adrian Thomas, Barlow Soper, and Lamar Wilkinson. 2005. Normal Personality Variables and Their Relationship to Psychological Reactance. *Individual Differences Research* 3, 2 (2005).
- [146] Gwendolyn Seidman. 2013. Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and individual differences* 54, 3 (2013), 402–407.
- [147] Jasjeet S Sekhon. 2009. Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science* 12 (2009), 487–508.
- [148] Jonathan A Shaffer, Andrew Li, and Jessica Bagger. 2015. A moderated mediation model of personality, self-monitoring, and OCB. *Human Performance* 28, 2 (2015).
- [149] Martin Shelton, Katherine Lo, and Bonnie Nardi. 2015. Online media forums as separate social lives: A qualitative study of disclosure within and beyond Reddit. *ICConference 2015 Proceedings* (2015).
- [150] Walter C Shipley. 2009. *Shipley-2: manual*. WPS.

- [151] Ian D Smith, Steven J Coombs, et al. 2003. The Hawthorne Effect: is it a help or a hindrance in social science research? *Change (Sydney, NSW)* 6, 1 (2003), 97.
- [152] Mark Snyder. 1979. Self-monitoring processes. In *Advances in experimental social psychology*. Vol. 12. Elsevier, 85–128.
- [153] 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.
- [154] Charles D Spielberger, Fernando Gonzalez-Reigosa, Angel Martinez-Urrutia, Luiz FS Natalicio, and Diana S Natalicio. 2017. The state-trait anxiety inventory. *Revista Interamericana Journal of Psychology* (2017).
- [155] Daniel R Stalder. 2007. Need for closure, the big five, and public self-consciousness. *The Journal of social psychology* 147, 1 (2007), 91–94.
- [156] Christina Steindl, Eva Jonas, Sandra Sittenthaler, Eva Traut-Mattausch, and Jeff Greenberg. 2015. Understanding psychological reactance. *Zeitschrift für Psychologie* (2015).
- [157] KP Suresh and S Chandrashekara. 2012. Sample size estimation and power analysis for clinical research studies. *Journal of human reproductive sciences* 5, 1 (2012), 7.
- [158] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [159] 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.
- [160] Harry C Triandis. 1977. Subjective culture and interpersonal relations across cultures. *Annals of the New York Academy of Sciences* 285, 1 (1977).
- [161] 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.
- [162] Zeynep Tufekci. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *ICWSM*.
- [163] Hanna M Wallach, Jain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning*. 1105–1112.
- [164] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. “I regretted the minute I pressed share”: a qualitative study of regrets on Facebook. In *Proceedings of the seventh symposium on usable privacy and security*. 1–16.
- [165] David Watson and Lee Anna Clark. 1999. The PANAS-X: Manual for the positive and negative affect schedule-expanded form. (1999).
- [166] Robert E Wilson, Samuel D Gosling, and Lindsay T Graham. 2012. A review of Facebook research in the social sciences. *Perspectives on psychological science* 7, 3 (2012), 203–220.
- [167] Pamela Wisniewski, Heather Lipford, and David Wilson. 2012. Fighting for my space: Coping mechanisms for SNS boundary regulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 609–618.
- [168] Sang Eun Woo, Andrew T Jebb, Louis Tay, and Scott Parrigon. 2018. Putting the “person” in the center: Review and synthesis of person-centered approaches and methods in organizational science. *Organizational Research Methods* 21, 4 (2018), 814–845.
- [169] Yunhao Yuan, Koustuv Saha, Barbara Keller, Erkki Tapio Isometsä, and Talayeh Aledavood. 2023. Mental Health Coping Stories on Social Media: A Causal-Inference Study of Papageno Effect. In *Proceedings of the ACM Web Conference 2023*. 2677–2685.
- [170] Dieter Zapf, Christian Dormann, and Michael Frese. 1996. Longitudinal studies in organizational stress research: a review of the literature with reference to methodological issues. *Journal of occupational health psychology* 1, 2 (1996).
- [171] Hui Zhang, Munmun De Choudhury, and Jonathan Grudin. 2014. Creepy but inevitable?: the evolution of social networking. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM.