Computational and Causal Examinations of Wellbeing in Situated Contexts by Leveraging Social Media and Multimodal Data

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

ABSTRACT

Our social lives are often embedded in our situated communities. Given that our experiences are not isolated, understanding how we cope with psychological and cognitive demands is essential for both individual and collective wellbeing. Wellbeing is typically assessed using surveys, which though accurate in snapshots, suffer from recall bias, are reactive, and do not scale. These limitations are surmountable by social and ubiquitous technologies. This dissertation uses social media in concert with multimodal sensing, which facilitate analyzing dense and longitudinal behavior at scale. By adopting machine learning, natural language, and causal inference analysis, this work infers wellbeing focusing on situated communities, such as college campuses and workplaces.

Before incorporating these assessments in practice, we need to account for confounds impacting behavior change. One such confound is the phenomenon of "observer effect" — that individuals may self-alter and deviate from their otherwise normal behavior because of the awareness of being "monitored". My proposed work studies this problem on social media behavior. On a multimodal sensing study of ~750 participants, I intend to conduct a causal study that adopts a theory-driven approach to model behavior change and measures how much and how long individuals are likely to modulate their behavior during study participation. Theoretically, this work will provide insights of this phenomenon in terms of both social media and longitudinal human behavior. Drawing on these insights, I expect to provide recommendations of correcting biases due to observer effect in social media sensing for human behavior and wellbeing. Broadly, my dissertation bears design and technological implications for social computing systems and various stakeholders to support wellbeing and in situated communities.

CSCW '20 Companion, October 17–21, 2020, Virtual Event, USA
 The definitive Version of Record was published in *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '20 Companion), October 17–21, 2020, Virtual Event, USA*, https://doi.org/10.1145/3406865.3418367.

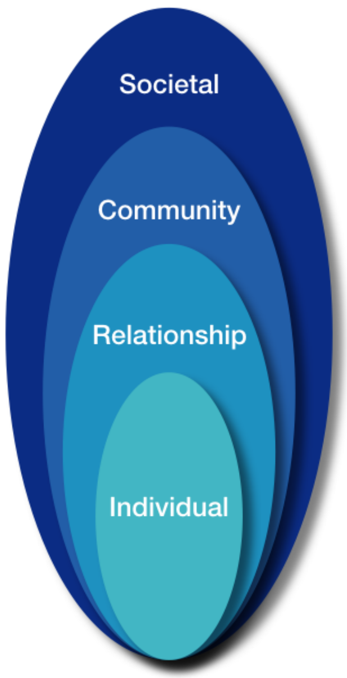

Figure 1: Social Ecological Model: Human behaviors can be considered to be deeply embedded in the complex interplay between an individual, their relationships, their communities, and societal factors. Social media provides a passive way to gather quantifiable signals about the social ecological dimensions relating to an individual's behavior [4].

CCS CONCEPTS

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

KEYWORDS

social media; wellbeing; situated; college campus; workplace; observer effect; causal inference

ACM Reference Format:

Koustuv Saha. 2020. Computational and Causal Examinations of Wellbeing in Situated Contexts by Leveraging Social Media and Multimodal Data. In *Companion Publication of the 2020 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '20 Companion), October 17–21, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3406865.3418367

INTRODUCTION

A core aspect of our social lives involves interactions with the communities we are situated in, such as our workplaces, residential neighborhoods and localities, school and college campuses, or even physically co-located interest communities, including third places [13]. The inter-connectedness and inter-dependencies of our interactions, experiences, and concerns, make individual and collective wellbeing interlinked in situated communities. For example, a nearby crime or violence can cause alertness and anxiety among several neighborhood residents. A better understanding of psychosocial dynamics can help devise strategies to address wellbeing concerns in situated communities.

Current methodologies to assess wellbeing suffer from limitations of scale and timeliness. On the other hand, social media, for its ubiquity and widespread use, can be considered as a "passive sensor" that can act as a complementary source of unobtrusive, real-time, and naturalistic data to infer wellbeing. Further, human behavior is a complex function of social, psychological, and environmental underpinnings which, when studied without accounting for confounds, may lead to unreliable and inconclusive findings. **By proposing computational and causal approaches that minimize the confounds, my dissertation leverages social media in concert with multimodal data to examine wellbeing in situated contexts.** However, the feasibility of proactive and real-time social media technologies for wellbeing may be sensitive to further confounds in practice, ones invisible in research using retrospectively collected data. **My dissertation makes a case for one such concern, "observer effect" and proposes to examine its pervasiveness in social media behavior.** This article briefly describes the motivations and contributions of my dissertation.

Studies of human behavior and wellbeing have typically relied on self-reported surveys. These approaches suffer from subjective assessments, recall and hindsight biases, and are often retrospective—information is gathered after an event or change has occurred [27]. Recent research values in-the-moment data recording and acquisition approaches, such as ecological momentary assessments (EMAs) capturing an individual's momentary state [25]. However, these approaches are challenged with scale, access, and cost [25]. EMAs often induce a response burden on participants through disruptions. Subsequently, researchers have employed various forms of passive sensing that facilitate

Situated communities are typically defined over geographic spaces where individuals share some form of physical colocation (same floor, building, locality, campus, etc.). Here, individuals share social interactions, bear common and distinctive social ties and interests, and access common resources and institutions dedicated to support and prosperity of community members [15]. The absence of appropriate and proactive support strategies may exacerbate both individual and collective wellbeing manifold due to the inter-dependencies and inter-connectedness in situated communities. For instance, lack of timely supportive interventions following an external crisis can proliferate community-cascading stress experiences leading to several negative consequences, such as post-traumatic stress disorder and acute stress disorder. However, capturing subjective aspects of individual lives in their situated contexts is a challenging undertaking. As already described, most existing approaches suffer from limitations, are reactive and largely based on discrete occurrences of events, and there is no way to continually and comprehensively assess wellbeing dynamics in situated communities.

unobtrusive data for studying human behavior [2, 6, 28]. Because social media data is recorded in the present by an individual, it also serves as a complementary *verbal* sensor to understand the psychosocial dynamics of an individual, beyond non-verbal passive sensors.

The potential of social media data for understanding complex human behaviors and wellbeing is explained by the *social ecological model* (Figure 1) [4], which posits that human behaviors and experiences of wellbeing are not isolated or individualized attributes, these have social underpinnings, and are impacted by events and factors in our relationships, communities, and societies. To obtain a better understanding of wellbeing, we need to include situated contexts, **and my dissertation focuses on *situated communities* as examples to consider situated contexts.**

My dissertation aims to overcome the gap of studying wellbeing in situated communities by using and complementing social media with multimodal data. I focus on two situated communities, *college campuses* and *workplaces*. I examine problems that are critical for situated communities. For instance, for college campuses, I study the effect of gun violence events on stress levels of college students [21] and the effectiveness of post-crisis interventions, particularly public service announcements on counseling recommendations following student deaths on college campuses [24]. For workplaces, I study how collective workplace dynamics such as organizational culture [7] or individualistic role dynamics [22] influence individual wellbeing and performance. The **contributions and novelties** of my dissertation are three-fold, described in the following three paragraphs:

I use social media data that particularly and uniquely capture the behavior of situated communities. My work leverages social media data that reflect online analog of the offline (physically co-located) situated communities. For example, I use college subreddit data for college campuses where college students express and share topics and interests about their day-to-day academic, personal, and college lives [20, 21, 24]. Similarly, I use Glassdoor data for workplaces, where workplace employees publicly express their workplace experiences [7]. These datasets uniquely allow us to capture the social and environmental context required for a better understanding of wellbeing.

I adopt theory-driven computational and causal methods to make conclusive research claims on wellbeing. We note that human behavior is influenced by several intrinsic and extrinsic factors in both normalcy and crisis. I adopt causal inference and computational approaches drawing on machine learning, natural language analysis, and statistical modeling. Causal methods minimize confounding factors and lead to stronger claims about cause-and-effect relationships regarding people's reactions to events or environments. For example, to understand the effect of gun-violence events on student stress, I account for stress attributable to academic, personal, relationship, environmental, and other factors in students' lives [21].

Despite its potentials, social media data comes with challenges. **I propose approaches of combining complementary multimodal data to overcome challenges of using social media data for understanding wellbeing.** In particular, I address two primary challenges, the lack of ground-truth [19] and the lack of social media presence altogether [16], by augmenting social media with complementary EMA and passive sensing data.

The validity and in-practice reliability of human-centered technologies suffer due to the unpredictability and complexity in human behavior along-with unaccounted confounds [14]. Individuals tend to change their use of social media over time, and factors such as platform design, audience control, social-desirability, and privacy concerns often lead people to self-present themselves differently than how they would have without this social context [5]. Further, in real-world scenarios, observer effect may additionally impact an individual's self-presentation. The ecological validity of these technologies remain unattested as observer effect is not typically accounted for. Literature posits that people may alter their behavior with the awareness of being monitored or observed. This phenomenon is known as the "observer effect", or popularly as the "Hawthorne effect' (see below), which has been cited to affect the reliability of study observations [11].

"The Hawthorne effect concerns research participation, the consequent awareness of being studied, and possible impact on behavior." — McCambridge et al. (2014) [11]

By situating the findings in an inter-disciplinary context including psychology and social science, my dissertation bears implications from theoretical, practical, methodological, and ethical perspectives catering to a variety of stakeholders including researchers and practitioners. A major implication concerns building tools and applications that leverage these data-driven methodologies to improve wellbeing in practice. For instance, campus and workplace welfare staff can use these tools to continually assess people's wellbeing and satisfaction, and act proactively with timely and tailored interventions. However, we understand that in reality, prospective data collection and use may be different from people's naturalistic course of action. A major reason is "observer effect", or an individual's tendency to modulate and change their behavior with the awareness of being observed [11]. My proposed work will provide a causal methodology to assess the presence and degree of observer effect in social media behavior, and will provide insights and recommendations on accounting for this behavior in prospective use of multimodal data that uses social media as a wellbeing sensor.

PROPOSED WORK: OBSERVER EFFECT

The same social ecological model that explains the potential of social media as a viable sensor in human-centered studies of wellbeing, also points out a caveat — the observers, who are also a part of a subject's ecology, may affect the subject's behavior (or the observer effect). I aim to study observer effect in social media behavior. I note that my completed work draws insights on data that is observational and *retrospectively* collected. These studies provide us the knowledge to build technologies of proactive wellbeing improvement. However, before diving into that, we should ask how these algorithms would perform if we conduct *prospective* data collection? Despite the potential, the validity of adopting computational social science on big data has been critiqued regarding its applicability in real-world [3, 9] (see left for more description).

In fact, social media behavior can be vulnerable to observer effect. Social media behavior can be considered to be an intentional and conscious behavior which an individual may alter if they feel "observed". To clarify, although social media data is a promising passive sensor of wellbeing, it is dependent on and sensitive to the active curation of individuals, and is characterized by their self-initiated desire to post or engage on a social media platform. *Observer effect remains a basic unexplored phenomenon that may bias observations regarding human behavior and wellbeing.*

Social media data also serves as an appropriate means to examine observer effect in longitudinal studies. Because social media data allows access to historical behavior of participants, i.e., participants' behavior before their study enrollment, allowing us an opportunity to model their normative behavior when they were supposedly "not-observed". I believe that a better understanding of observer effect in social media sensing would not only make researchers aware of what they can expect but also can make us think towards approaches to correct and account for this effect in study designs.

Objective and Expected Contributions. I propose to examine observer effect on social media behavior. My examination is situated within a year-long multisensor study where participants consented to their social media data. To minimize the effects due to confounding factors, I intend to use causal inference methods. I will adopt a theory-driven approach to operationalize social media

The data for my proposed work includes year-long data for U.S. information workers who participated in a multisensor study of wellbeing, Tesserae [10, 12, 18]. Table 1 summarizes the demographics and individual differences of those who provided us Facebook data. We find that the data is fairly well distributed across different psychological traits.

Table 1: Summary of demographics and individual differences in the participants whose data will be studied for observer effect.

Covariates	Values / Distribution
Demographic Characteristics	
Gender	M|F
Age	(21:64)
Executive Function (Shipley)	
Fluid	(5:24)
Crystallized	(0.0:40.0)
Personality Trait (BFI)	
Openness	(2.1:5.0)
Consct.	(1.9:5.0)
Extraversion	(1.7:5.0)
Agreeabln.	(2.1:5.0)
Neuroticism	(1.0:4.6)
Affect (PANAS) and Anxiety (STAI-Trait)	
P. Affect	(19.0:49.0)
N. Affect	(10.0:40.0)
Anxiety	(20.0:67.0)

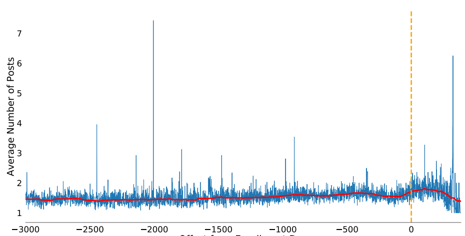

Figure 2: Prelim. Analysis Example: Average number of posts per day on relative offset from day of enrollment (day 0).

Acknowledgement. I thank my advisor, Munmun De Choudhury.

behavior of individuals, and examine if and how their behavior deviated from their expected behavior post their study participation or the awareness of being "sensed". From a methodological perspective, my proposed work will contribute a causal-framework of modeling and inferring observer effect in human-centered studies with social media sensing. My expected contributions also include providing insights regarding whether observer effect occurs, how long does it last, and how do its occurrences vary across participants. Drawing on these insights, the implications of my work will recommend strategies to correct biases due to observer effect in social media sensing studies.

Data and Methods

To measure observer effect, I will measure the deviation from normative behavior of an individual *caused* by the presence of observer. However, behavior changes could be because of several reasons, and to delineate the effects of observer from other kinds of confounders, I will adopt a causal approach. Typically, causal studies involves a treatment, and examining time-series changes before and after the treatment can be conducted using an interrupted time series (ITS) approach. I will operationalize *treatment* as study enrollment. Because my dataset does not consist of any untreated (or control) individual, I will obtain counterfactual dataset based on synthetic control approaches [1]. I will build machine learning models learning people's historical social media behavior to predict future social media behavior by applying time-series forecasting techniques. This forecasted data can function as a user's counterfactual behavior with some error margin, i.e., the behavior they might have shown if they did not enroll in the study. If other confounders are minimized, the difference in actual and counterfactual behavior corresponds to the user's deviation in behavior due to observer effect.

Observer effect is known to affect an individual's expressive behavior. I will quantify expressive behavior as affective, behavior, and cognitive attributes on social media [17, 23, 24]. I will borrow from prior work on operationalizing intimacy of self-disclosure in terms of breadth and depth of disclosure [8]. I will use computational linguistic methodologies like psycholinguistic characterization, word embeddings, and topic modeling, and adopt time-series and predictive modeling, including accounting for trend and seasonality. After measuring the pervasiveness of observer effect in my study, I also intend to contextualize my findings with prior literature to understand who are most likely to show this behavior and for how long [26]. For this purpose, I will take a deeper dive and introspect the relationship of observer effect with individual differences (such as personality traits and executive function). To understand these relationships, I will adopt regression modeling. Finally, I will situate these observations with the theory and provide insights regarding factors constituting a greater propensity and duration of observer effect.

REFERENCES

[1] Sebastian Bauhoff. 2014. The effect of school district nutrition policies on dietary intake and overweight: a synthetic control approach. *Economics & Human Biology* 12 (2014).

[2] Mehrab Bin Morshed, Koustuv Saha, Richard Li, Sidney K. D'Mello, Munmun De Choudhury, Gregory D. Abowd, and Thomas Plötz. 2019. Prediction of Mood Instability with Passive Sensing. *PACM IMWUT* (2019).

[3] danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* (2012).

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

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

[6] Vedant Das Swain et al. 2019. A Multisensor Person-Centered Approach to Understand the Role of Daily Activities in Job Performance with Organizational Personas. *PACM IMWUT* (2019).

[7] Vedant Das Swain, Koustuv Saha, Manikanta D Reddy, Hemang Rajvanshy, Gregory D Abowd, and Munmun De Choudhury. 2020. Modeling Organizational Culture with Workplace Experiences Shared on Glassdoor. In *CHI*. ACM.

[8] Sindhu Kiranmai Ernala et al. 2018. Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. In *ICWSM*.

[9] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The parable of Google Flu: traps in big data analysis. *Science* 343, 6176 (2014), 1203–1205.

[10] Stephen M. Mattingly et al. 2019. The Tesserae Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. In *CHI Ext. Abstracts*.

[11] 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.

[12] Shayan Mirjafari et al. 2019. Differentiating Higher and Lower Job Performers in the Workplace Using Mobile Sensing. *Proc. ACM IMWUT* (2019).

[13] Ray Oldenburg. 1999. *The great good place: Cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community.* Da Capo Press.

[14] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* (2019).

[15] Dennis E Poplin. 1979. *Communities: A survey of theories and methods of research.* Macmillan.

[16] Koustuv Saha et al. 2019. Imputing Missing Social Media Data Stream in Multisensor Studies of Human Behavior. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*.

[17] Koustuv Saha et al. 2019. The language of lgbtq+ minority stress experiences on social media. *PACM HCI* CSCW (2019).

[18] Koustuv Saha et al. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In *CHI Ext. Abstracts*.

[19] Koustuv Saha, Larry Chan, Kaya de Barbaro, Gregory D Abowd, and Munmun De Choudhury. 2017. Inferring mood instability on social media by leveraging ecological momentary assessments. *IMWUT* (2017).

[20] Koustuv Saha, Eshwar Chandrasekharan, and Munmun De Choudhury. 2019. Prevalence and psychological effects of hateful speech in online college communities. In *WebSci*.

[21] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *PACM HCI* CSCW (2017).

[22] Koustuv Saha, Manikanta D Reddy, Stephen M Mattingly, Edward Moskal, Anusha Sirigiri, and Munmun De Choudhury. 2019. LibRA: On LinkedIn based Role Ambiguity and Its Relationship with Wellbeing and Job Performance. *PACM HCI* CSCW (2019).

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

[24] 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*.

[25] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. 2009. Experience sampling: promises and pitfalls, strength and weaknesses. In *Assessing well-being*. Springer, 157–180.

[26] Mark Snyder. 1979. Self-monitoring processes. In *Advances in experimental social psychology*. Vol. 12. Elsevier, 85–128.

[27] Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response*.

[28] Rui Wang et al. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Ubicomp*.