PASSIVE SENSING FRAMEWORKS FOR THE FUTURE OF INFORMATION WORKERS

A Dissertation
Presented to
The Academic Faculty

By

Vedant Das Swain

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing
College of Computing

Georgia Institute of Technology

August 2023

PASSIVE SENSING FRAMEWORKS FOR THE FUTURE OF INFORMATION WORKERS

Thesis committee:

Dr. Munmun De Choudhury
School of Interactive Computing
Georgia Institute of Technology

Dr. Gregory D. Abowd
School of Interactive Computing
Georgia Institute of Technology
College of Engineering
Northeastern University

Dr. Sauvik Das
School of Interactive Computing
Georgia Institute of Technology
Human-Computer Interaction Institute
Carnegie Mellon University

Dr. Thomas Plötz
School of Interactive Computing
Georgia Institute of Technology

Dr. Anind K. Dey
Information School
University of Washington

Dr. Shamsi T. Iqbal
Viva Insights
Microsoft Research Redmond

Date approved: July 19, 2023

I have found that it is the small everyday deed of ordinary folks that keep the darkness at bay. Small acts of kindness and love.

Gandalf

To Ma and Papa, who encouraged me to start.

To Kate, who strengthened me to keep going.

ACKNOWLEDGMENTS

Before I applied for PhD programs I had chanced upon something rare. Research labs with happy students, thriving with passion. Frankly, I did not know how rare they were untill I set out on my own pursuit. As I went through the program I learned how much of a life is lived outside of the pages of research we publish. A lot is probably lived outside the pages we do not. As scientists the only thing we control is our study design. Everything outside it, not so much. Life happens. When I look back at my own PhD journey, the first word that comes to my mind is "unremarkable" and I mean that in the best way. The people around me protected my sails through every remarkable challenge that arose.

I am grateful to have two wonderful advisors, Gregory Abowd and Munmun De Choudhury. A best of both worlds combination. In *Star Wars* terminology, I learned the ways of the *force* from one and the ways of the *lightsaber* from another. Gregory, you were the first person who said I was "PhD material". To this date I do not know what you measured, but we made it. Munmun, you let me pursue questions I wanted and made sure I ran that last lap. You even humored every mundane question I had, from dress codes to salutations. Both of you have taught me a lot about research and even more about being a good researcher. Thank you for all the wisdom and all the banter.

I was fortunate to be guided by a stellar thesis committee, Thomas Plötz, Sauvik Das, Anind Dey, and Shamsi Iqbal. Thomas, I hope to one day emulate your level of organization, preparedness, and discipline. Sauvik, you nudged me to reconcile my values with my research and even value my research itself. Anind, your advice helped me view my research in a far more wholesome way than a specialized degree is designed to make you think. Shamsi, you showed me that my work is actually meaningful in the world outside academia. Outside of my committee I have been blessed with advice from Mary Czerwinski, Gloria Mark, Adam Fourney, and Javier Hernandez. All of you have counseled me beyond my research and I thank you for your mentorship.

A PhD is often described as lonely. Intellectually, it sometimes was. Besides that, though, my peers never let it be. I have smiled through the hardest deadlines because of these people. Dong Whi Yoo, you are my only *full academic sibling* and your role in my life has been much larger. Koustuv Saha, your optimism towards research has always made me find a way out, especially through your terrible defensive line in FIFA. Sindhu Ernala, you taught me to be better at taking compliments and I did eventually come across days where I held them tight. Youngwook Do, you showed me that no matter how dire the situation, we can always find a pun. Jiawei Zhou, you always reminded me the importance of caring for myself. Qiaosi Wang, you stayed patient with me through every up, down, and tangent I threw at you. My gratitude exceeds the names I have mentioned so far. Many of you have mattered through coffee breaks, shared meals, walks in hallways, and football matches. Thank you to *GT UbiComp group*, *SocWeB* lab, the interns from *MSR*, and other students beyond who shared this journey with me.

My research would have been a fragment of itself without the tireless dedication of my coauthors and assistants, Hemang Rajvanshy, Manikanta Reddy, Sonia Sargolzaei, Victor Chen, Lan Gao, William Wood, Jiajia Xie, Jingjing Ye, Siva Karthik Ramesh, and other contributors to *Tesserae* and *CampusLife*. Thank you for your support. I am waiting to see all of your path–breaking achievements to come.

I too was once an undergraduate student assisting in research. In India, it was rare to gain hands-on research experience as someone fresh out of school. At the time, it felt risky to complete my undergraduate studies from a lesser known and very young institute in India. Yes, IIIT-Delhi was nascent, but we were also ambitious. My vision for research was seeded thanks to my undergraduate advisors, Vinayak Naik, Amarjeet Singh, and Ponnurangam Kumaraguru. Thank you for the initial push that got me in this line.

Without my family, I would never find and cultivate the families I mentioned earlier. They have been with me unconditionally My mother has given every bit of herself to every responsibility she has, professional and personal. She let me learn the way I wanted, let

me watch TV when I liked but also got me big stacks of encyclopedias to browse through. Ma, I wish I could quantify how much those acts have helped me. Thank you for always feeding me when I was too stressed to even taste your meals. My father is the hardest working man I know. I might never be able to emulate his perseverance but I have learned to use his humor to lighten even the darkest circumstances. Papa, when I asked you what the return on investment is for a PhD, you told me that the return is my education. Thank you for always turning that extra light on when I studied. My brother was my original partner in crime. I aped everything he did but it always amazed me how deeply he studied something he got interested in. Bhai, your obsessive nerdiness is what really trained my research muscles. Thank you for never letting go of that. I am sorry for the times I have taken any of you for granted. Always remember, you mean everything.

Finally, I wanted to thank the person who saw me through the end. My wife, Kate. Remember when I said that during a PhD, "life happens"? Well, it did. I was not planning on it and I met a person who would give me so much that I would learn to live outside my research. Without Kate my wellbeing would have plummeted over the last three years of my PhD. Kate, your "sacred" weekends have not only helped me recenter my life, but have also given me perspective on why I work. You have not just been my best friend and confidant but been anyone I needed you to be. You have given me feedback on presentations, styled me for my talks, and even helped me recruit participants. You could probably write my dissertation if I had asked. Might have turned out better. Your impact on my PhD cannot be justified with words. I will never be able to articulate the hilarity of our conversations, nor will I be able to describe the delicious taste of your cooking. I know I have a lot more research to do ahead. It is frightening but we will face it. I was privileged through my journey so far and thanks to you, Kate, I always will be. Time to find out what we achieve with more successful weekends and long walks through the park.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ABM Our approaches to use mobility information from WiFi to design policies with Agent–Based Modeling

EN Approaches that use course Enrollment data to design closure policies

WIMOB WiFi Mobility Models

LC Localized Closure policies

RI Remote Instruction policies

S1 Behavioral Scenario: Persistent

S2 Behavioral Scenario: Non-Residential Avoidance

S3 Behavioral Scenario: Complete Avoidance

AP Access-Point

COVID-19 Coronavirus Disease 2019

CSCW Computer Supported Collaborative Work

CWB Counterproductive Work Behaviors

FFM Five Factor Model

GT Georgia Institute of Technology

HCI Human-Computer Interaction

HRM Human Resource Management

I/O Industrial–Organizational Psychology

ICT Information and Communication Technologies

IRB In-Role Behavior

ITP Individual Task Proficiency

IW Information Workers

NPI Non-Pharmaceutical Intervention

OC Organizational Culture

OCB Organizational Citizenship Behaviors

P–O Fit Person-Org Fit

PCA Principal Component Analysis

PSAI Passive Sensing–enabled AI

SEIR Susceptible–Exposed–Infectious–Removed Template for Simulating Infectious Disease

SMAPE Symmetric Mean Absolute Percentage Error

SVM Support Vector Machines

TAM Technology Adoption Model

SUMMARY

Work sustains our livelihoods and is integral to a fulfilling life. As workers, our day-to-day experiences determine our performance and our wellbeing. Yet, the behavioral richness of a worker's everyday life is often overlooked by traditional survey methods used to describe work experiences and outcomes. **I posit that we can gain a more naturalistic understanding of work by leveraging everyday technology to develop passive—automatic and unobtrusive— sensing applications with machine learning.**

My research investigates the potential of multimodal sensing to computationally model worker behaviors and related psychosocial outcomes. I rely on passively sensing naturalistic behaviors with technology readily available in information work; including wearables, mobiles, desktops, Bluetooth beacons, WiFi router networks, and social media. For individual workers, my analyses thus far have revealed clusters of mutable day-level activities (e.g., batched phone use and better sleep) that explain improved job performance above and beyond inflexible personality assessments. For groups, I have leveraged coarse signals from WiFi routers to empirically demonstrate that collocation patterns of team members indicate their performance. Additionally, I have used observations from Bluetooth beacons to compute eigenvectors of desk break routines and validate a new behavior–based construct to measure how workers fit within a group. Further, for an entire community of workers, I have shown that by repurposing anonymized and archival data we can provide a flexible toolkit that can inform large-scale policy change. To support resumption of work during epidemics, I have expressed campus mobility with WiFi router logs to design response infection control strategies that minimize disruption to learning, social contact, and space utilization. Similarly, to describe normative perceptions of worker communities, I have described organizational culture from self-initiated posts on social media.

Besides uncovering the utility of passive sensing frameworks, my research also critically reflects on the challenges of deploying these frameworks in the real world. My re-

search has investigated the semantic gap in the predictive power of different sensing streams and furnishes computational evidence to guide better sensor deployment practices to infer worker wellbeing. I also engage with information workers to highlight the expected norms of being data subjects of passive sensing enabled AI. This research highlights both empowering and exploitative visions of this technology from the eyes of the worker. Lastly, I explore the value of different design variations in making information flows for passive sensing enabled AI more acceptable to workers. Through these findings I describe the trade-offs between different passive sensing frameworks in terms of their types of sensors, scope of sensors, types of insights, and sharing of insights.

Collectively, my research not only has implications for predictive uses of passive sensing, but also informs domain experts (such as organizational psychologists and personnel management) to gather empirical evidence for workplace phenomena, develop novel measures indicative of work outcomes, and inform experiences of large workforces. The aim of my research is not to make the workers work for these technologies, but to make these technologies work for the worker. This dissertation encourages methods to make passive sensing for workers more holistic, accurate, and humane.

CHAPTER 1
INTRODUCTION

We seek out work where we can provide value consistently without compromising our needs. In manufacturing work, value was related to efficiency in production. A stakeholder might ask, ”are they doing things right?”. For a weaver, the number of correctly woven baskets in a given day could indicate this. Thus, value was expressed in terms of *productivity* both as a process and an outcome [1]. However, measuring value becomes challenging when the processes and outcomes are intangible or ambiguous. This is particularly the case when a worker's primary role is to manipulate information, known as information work [2]. Now, a stakeholder might also ask, ”are they doing the right things?”. Thus, we move on from productivity to broader concepts of *performance* [1] and *wellbeing* [3] to express effective workers. For an analyst, the number of completed issues in a day is not a sufficient indicator. We need to consider the quality of the solutions, the severity of the issues, their accountability towards supervisors, their availability to peers, their flexibility towards requirements and many other aspects [1]. Moreover, how is this worker achieving these outcomes? We need to account for their happiness, stress, overall satisfaction at work, experience outside work and several other related factors [3]. Traditionally, these aspects have been studied with survey–based methods which often overlook the richness of a worker's everyday life. **I posit that we can gain a more naturalistic understanding of workers by leveraging everyday technology dispersed in their ecology to develop passive—automatic and unobtrusive— methods to study worker experiences.**

As noted above, performance and wellbeing are important, yet complex to ascertain simply by observer reports [4, 5]. To tackle this, researchers in I/O have developed several survey instruments for information workers to self-report their experiences. Accordingly, personnel management and human resources have followed suit and started relying on these instruments and what they measure. However, these surveys are often one-time measures [6] that are scarcely revisited and too burdensome to complete for large set of workers multiple times. As a result, such instruments ignore changes and dynamic developments in a worker's life. In fact, the constructs measured are often rigid, such as personality [7]. The reductive nature of such measures can encourage selection and rejection of workers as opposed to improving individual workers to their strengths. Also, for what surveys capture, the responses are vulnerable to biases. Respondents tend indulge in impression management and selective disclosure [8, 9]. Several biases are accentuated in the work environment due to fear of the consequences [10, 11]. Therefore, surveys have significant methodological limitations in expressing a worker's performance and wellbeing. Moreover, we know from the Social-Ecological Model [12], that a worker's experience is not limited to the self, but also their team and larger organizational variables. Survey based instruments fail to consistently reflect this complex interplay of factors that relate to a worker's effectiveness. Instead, we need to investigate new approaches that can automatically and unobtrusively explain information workers' experiences in a naturalistic setting. To meet this need, I turn to passive sensing — a dynamic approach to learning about people without any active effort from the user [13].

Today, digital technologies are integral to an information workers' routine. They regularly use with devices like personal computers and smartphones. Additionally, they rely on internet connectivity and communicate over social media. Everyday, workers interact with these technologies and imprint some traces of their behaviors onto these.

Everyday digital technologies can be repurposed as an ecological lens to describe the performance and wellbeing of information workers.

Through my dissertion, the first set of research questions I discuss investigate the feasibility of these technologies to provide meaningful insights of workplace effectiveness. I take a holistic view of the information worker by incorporating their ecological context and demonstrate a variety of passive sensing frameworks by answering the following questions.

RQ I. *How can passive sensing frameworks explain individual worker outcomes?*

RQ II. *How can passive sensing frameworks explain social dynamics at a workplace?*

RQ III. *How can passive sensing frameworks inform organizational change?*

My research harnesses multimodal sensing to computationally model worker behaviors and related psychosocial outcomes. I rely on passively sensing naturalistic behaviors with technology readily available in information work; including wearables, mobiles, desktops, Bluetooth beacons, WiFi router networks, and social media. For individual workers, my analyses thus far have revealed clusters of mutable day-level activities (e.g., batched phone use and better sleep) that explain improved job performance above and beyond inflexible personality assessments. For groups, I have leveraged coarse signals from WiFi routers to empirically demonstrate that collocation patterns of team members indicate their performance. Additionally, I have used observations from Bluetooth beacons to compute eigenvectors of desk break routines and validate a new behavior–based construct to measure how workers fit within a group. Further, for an entire community of workers, I have shown that by repurposing anonymized and archival data we can provide a flexible toolkit that can inform large-scale policy change. To support resumption of work during epidemics, I have expressed campus mobility with WiFi router logs to design response infection control strategies that minimize disruption to learning, social contact, and space utilization. Similarly, to describe normative perceptions of worker communities, I have described organizational culture from self-initiated posts on social media.

Despite the potential uses of passive sensing frameworks for work, it must not be confused with a silver bullet. Like surveys before it, through my studies I observed the lacking in this new approach. The remaining part of the dissertation is focused on self-reflective questions. This segment of my dissertation begins with an inquiry of practical constraints in these frameworks by answering the following question.

RQ IV. *What are the methodological challenges of building effective passive sensing frameworks for information worker experiences?*

Thus, I critiqued the predictive power of different sensing streams and furnish computational evidence to guide better sensor deployment practices to infer worker wellbeing. As a result, this dissertation not only has implications for predictive uses of passive sensing, but also informs domain experts (such as organizational psychologists and personnel management) to gather empirical evidence for workplace phenomena, develop novel measures indicative of work outcomes, and inform experiences of a large workforce.

Answering RQ I–IV adds to the growing evidence in the field that passive sensing frameworks promise granular data-driven insights that can inform new ways to work without any burden of manual reporting. However, despite the negligible effort, adoption of these technologies is hindered with evolving worker concerns surrounding centralization of data with organizations, an amorphous work-home boundary introduced by burgeoning shifts to remote work paradigms, and the power asymmetry in information work settings. Therefore, to complete my dissertation, I addressed the following question.

RQ V. *What are the sociotechnical challenges of deploying passive sensing frameworks in an information worker's ecosystem?*

I pursued this research question through a worker-centred lens to emphasize their voice as data subjects (or data providers) of such passive sensing frameworks. My findings were generated from worker opinions and perceptions through two studies. In the first, I conducted scenario–based interviews with information workers. This qualitative study brought to light the expected norms of appropriate sensing and information distribution. My findings express how workers imagine Passive Sensing–enabled AI (PSAI) would exist in their work and why it could lead to powerful and punitive applications. That study also serves as a formative counterpart to the next study. In the second study, I investigated the factors of passive sensing frameworks that indicate a worker's perception of its burdens and benefits.

I conducted an interactive experimental vignette-based online survey for information workers to express their perceived utility and harms for different scenarios of passive sensing. I find statistical evidence to reflect the trade-offs between different design decisions in implementing passive sensing for work. As a result, my study provides a roadmap to conceptualize more worker-centred passive sensing information flows. Overall, my dissertation is motivated to guide subsequent research in the domain of using ubiquitous technology to understand work and workers, and to provide implications that underscore the adoption of human-centered sensing frameworks towards the future of work.

Organization of the dissertation. This dissertation is organized as follows. chapter 2 describes the background of information work and variables of interest in domain literature. Additionally, it also describes prior work on passive sensing for work. chapter 3 describes the common data and materials that feature through the studies described in this paper. Particularly, I summarized two passive sensing datasets for information work. Chapters 4–8, contain various studies that answer the research questions that contribute to my thesis. In accordance with my multi-level approach, chapter 4 describes a study to explain individual performance with multi-modal sensing, chapter 5 describes two studies that present evidence for the sensing of interpersonal dynamics, and chapter 6 describes two studies that illustrate passive sensing frameworks that infer organizational characteristics. After these chapters, my dissertation takes a critical stance that reflects on the previous chapters and other related work. chapter 7 demonstrates the methodological challenges of passive sensing for mental wellbeing. chapter 8 investigates the socio-technical challenges of deploying passive sensing for information workers. Finally, chapter 9 discusses the implications of my research for responsibly designing passive sensing for worker wellbeing.

Table 1.1: Outline of dissertation research

Study	Theme	Summary	Sensor(s)	Index
Understanding the role of daily activities in job performance with organizational personas [14]	Passive sensing frameworks to explain individual worker outcomes	Exhibits passive sensing as an analytical lens for personnel management to understand performance beyond personality assessments	Smartphone, Wearable, Bluetooth Beacons	chapter 4
Leveraging WiFi Network Logs to Infer Collocation of Teams and its Relationship with Performance [15]	Passive sensing frameworks to explain group dynamics at a workplace	Utilizes passive sensing to empirically validate the social phenomena of spatiality, which is related to performance of coworkers	WiFi APs	section 5.1
A Study of Person–Organization Fit Through Latent Activity Routines [16]	Passive sensing frameworks to explain group dynamics at a workplace	Illustrates passive sensing as a means to conceive new measures of team cohesion that influence the performance and wellbeing of workers	Smartphone, Bluetooth Beacons	section 5.2
Characterizing Organizational Culture with Passively Collected Accounts of Workplace Experiences [17]	Passive sensing frameworks to inform organizational decisions	Demonstrates repurposing publicly accessible unstructured language data to gauge organizational culture and explain performance	Social Media	section 6.1
Modeling Organizational Networks to Aid Infectious Disease Crisis Response [18]	Passive sensing frameworks to inform organizational decisions	Demonstrates repurposing of anonymized aggregated data to describe community behavior can help organizations resume operations while containing a crisis	WiFi APs	section 6.2
Semantic Gap in Predicting Mental Wellbeing through Passive Sensing [19]	Methodological challenges in passively inferring psycho-social experiences of workers	Presents empirical evidence to critiques contemporary practices of multimodal passive sensing predictions for wellbeing	Smartphone, Wearable, Bluetooth Beacons, Social Media	chapter 7
Worker Perspectives on Passive Sensing Enabled AI Phenotyping of Performance and Wellbeing [20]	Contextual norms to inform deployment of passive sensing into an information work settings	Highlights the expectations workers have for appropriate sensing and reasonable information sharing paradigms within the power dynamics of work	Various	section 8.1
Vignette Analysis of Acceptable Passive Sensing Enabled AI Phenotyping for Workers	Components of passive sensing frameworks that determine adoption	Identifies factors that distinguish passive sensing information flow variations in terms of perceived utility and perceived harm	Various	section 8.2

CHAPTER 2
BACKGROUND

2.1 Information Work

The working population is composed of a variety of different tasks that can be bucketed into different sectors. In general, this dissertation collects studies that utilize passive sensing frameworks to clarify behaviors of *information workers* and describes methods to support them. Information workers primarily provide value to an organization by gathering, interpreting or creating information to support enterprise decisions [2]. Such tasks are distinct from work that involves manual labor or needs physical proficiency. This includes farmers, factory workers, and builders — known today as blue–collar work. In fact, manual work has dominated human civilization all the way through the 20^{th} century. More recently, however, information work has emerged to play a significant role in shaping how our economies function. In the 1966, Drucker first described such workers in such roles as *knowledge workers* [21]. Colloquially speaking, these workers "think for a living" [22] or white–collar workers. In the strictest sense, Drucker described these workers as specialists who create knowledge and our singular sources of it too [21]. However, with advancements in computing technology this knowledge was externalized into storage devices as information. Personal computers and later smartphones have made this information access ubiquitous. As a result, many non-specialist workers could understand, exchange and manipulate this information. Thus, I use the term "information work" throughout this dissertation as it is more inclusive of broad variety of work roles prevalent today. Moreover, their work is inherently tied to the digital technologies that I will harness to understand their work experience. In current verbiage, information workers and knowledge workers are often used interchangeably. Several studies in the Human-Computer Interaction (HCI) and adjacent fields have studied information workers and their use of Information and Communication Technologies (ICT) [23, 24, 25, 26, 27].

At the time of writing this, most estimates of report at least a billion information workers across the globe [28, 29]. This includes analysts, managers, programmers, engineers, accountants, lawyers, and academics. A large part of this boom can be attributed to the increased penetration of ICTs that are fundamental for these workers to achieve their tasks. While these workers are critical to the economy, the work force includes other kinds of workers too. However, I am particularly interested in information workers because it has been historically challenging to identify what it means for them to work effectively [30]. Especially because thinking cannot be evaluated the same way productivity in manual work would be measured, i.e., the production of more artifacts [31]. My thesis proposes that the very technology that information workers interact with regularly can be used as a lens to describe their effectiveness at work.

2.2 Traditional Indicators of Worker Effectiveness

In manual work, a positive indicator is when a worker is *doing things right*. A worker's performance is typically defined by the efficiency of their output, or how quickly they can produce tasks. By contrast, in information work, assessments of performance need to consider if the worker is *doing the right things* for the organization [30]. Thus, we shift a focus from efficiency, to effectiveness [1] — a more qualitative outlook towards a worker's outcomes. Rotundo and Sackett define job performance as, ”those actions and behaviors that are under the control of the individual and contribute to the goals of the organization”. An information worker's effectiveness is not only determined by their output but their overall wellbeing that informs sustainable and enriching work experiences. An information worker's effectiveness can be explained by a variety of factors. The domain literature on these indicators inspire my research and drive my investigations with passive sensing frameworks.

Taking a multi-level approach based on the Social-Ecological Model [12], my dissertation discusses indicators of worker effectiveness along different levels (moving outward from the worker): (i) intrinsic worker traits, (ii) their social relationships, and (iii) more encompassing norms within their organization.

2.2.1 Personality

A widely studied individual attribute that explains job performance is their personality. This is a characteristic set of motivations or perceptions that can describe how a person is predisposed to interact with their surroundings. In the context of work, an extensively studied framework to express personality traits is the *Five Factor Model* (FFM) [7]. As per FFM, a worker's personality can be measured along five dimensions; conscientiousness, neuroticism, extraversion, agreeableness, and openness. Each of these is uniquely related to different aspects of job performance.

Workers who are dutiful, organized and focused, are regarded to display high *conscientiousness*. High performing workers have typically also been highly conscientious [33, 34, 35, 36]. Performance is also related to a worker's ability to socialize and assert themselves in the company of others. FFM describes this trait as *extroversion*. Studies have shown extroversion to be correlated with high performance in people-facing roles and training proficiency [34, 37]. Since most information work is collaborative, people who are kind and helpful are often rated as high performers [34, 37]. These compassionate characteristics are operationalized as *agreeableness*. In roles requiring creativity and innovation, information workers are required to be intellectually curios, adventurous, and inquisitive [38, 39]. These attributes are reflected by the *openness* trait. Therefore, depending on the role specifics, workers that measure highly on the aforementioned traits are considered to be high performing. On the other hand, a negative indicator of performance is a worker's tendency to be less emotionally stable and anxious. This propensity is known as *neuroticism* and has often found to be negatively correlated to job satisfaction [40].

A majority of studies exploring personality indicators of performance are based on one-time survey instruments that collect subjective self-reported data. Having said that, not all of this research has been entirely agnostic to the concept of context. Although personality has consistently provided a strong signal representing job performance, it is operationalized as a fairly stable construct that does not typically vary on a daily basis. It is restricted to what a person *is* and agnostic to what they *do*. Beyond a worker's intrinsic personality, Industrial–Organizational (I/O) Psychology studies have hypothesized situational factors that explain the variability in job performance [41]. This motivates my examination into everyday actions that indicate job performance and understanding how it can accompany the existing influence of personality on individual measures of performance. Since workers leave traces on technologies they interact with and interact around, my work aims to model these behavioral signals to explain their work experience.

2.2.2 Interpersonal Dynamics

Although personality can be a significant determinant of job performance, a worker's ability to work is also related to their larger social ecology [12]. For instance, when workers are collocated with their coworkers they find more opportunities for synchronous social interactions [42, 43] (e.g., open offices or adjacent cubicles). Working in the same space with the presence of others is not limited to verbal discussions and active sharing of resources. Even the presence of others working towards a common goal allows for subtle exchange of information through gestures and expressions [44] (e.g, is a teammate struggling, are they too absorbed or are they available for feedback). Additionally, collocation provides shared context that comprises common points of reference (e.g., whiteboards, post-it notes, or verbal concepts) [44]. Moreover, it supports informal interactions that can help "opportunistic information exchange" and improve social ties with teammates [44]. These social interactions keeps team members up-to-date, available for feedback and therefore agile and innovative [42, 45]. Social interactions while collocated in the same space can also improve social ties between members [46] and therefore improve performance [47]. Even subtle cues of collocated social interactions are associated with longer, continuous periods of focused activity [43]. However, very few studies have observed the importance of collocation patterns in-the-wild. Human observation cannot represent behaviors of multiple groups over time and gather evidence the importance of these interactions between workers. As a result, such influences are largely ignored when considering the outcomes of a worker. My work fills this gap by using passive sensing frameworks to highlight the importance of these interconnected behaviors within the workforce.

The social dynamics of the work place extend beyond the characteristics of events where coworkers are collocated. These dynamics also emerge from the larger process of socializing into the group. At work, the socialization of a worker, which is the interaction between their perspectives and that of the broader community, can largely dictate their work experiences [48]. I/O psychologists would describe this phenomenon as "person–organization fit"(P–O Fit) [49]. This idea of *fit* can explain both the satisfaction and tenure of employees [50, 51]. P–O Fit can be also manifest when when workers' expectations do not match that of their organization. For instance, inequity in pay can lead to attrition [52, 53], whereas a mismatch of job norms can imply reduced performance [54]. These differences in worker–organization perspectives are often studied in the domain of I/O through survey instruments. However, many of these approaches to study these differences in socialization are considered "reductionist" as they condense the organization's norms into a finite set of scales [55]. Other survey approaches expect participants to self-determine the differences on a single–occasion [56, 6]. Moreover, all these approaches tend to measure attitudes of workers, not the actual behaviors that describe socialization. I overcome these limitations by using passive sensing frameworks to develop new behavioral explanations of socialization at workplaces.

2.2.3 Organizational Culture and Policy

Although organizations comprise of multiple interconnected workers, an individual worker's experience can be determined by other workers who are not directly related to them. Decisions made from higher up, or even further away in a network of connections can shape how workers do their work.

One of the fundamental aspects that permeates to individual workers is *organizational culture* (OC). Formally, this refers to a socio-cognitive model of emergent standards and norms that help individuals to make sense of their surroundings [57, 58]. OC materializes from the interplay of top-down expectations (from management) and bottom-up norms (from present and former workers) [59]. OC can have stark affect on workers. For example, toxic or unethical cultures can deplete employee morale [60] and lead to turnover [61]. By contrast, an OC that is supportive and built on positive incentive leads to greater satisfaction [62] and reduction in misconduct [63]. Traditionally, OC has been studied using a variety of theoretical frameworks [50, 64] but these assessments are limited in organizational settings because of the power dynamics [65]. When such instruments are administered in a workplace, workers might feel uncomfortable sharing their opinion [10, 11] or fail to respond honestly [66]. My work extends these theoretical models by using passively aggregated worker perspectives that are self-initiated and candid.

Another encompassing determinant of worker effectiveness is workplace policy. Especially during times of crisis, organizations must take drastic decisions that impact their entire workforce. For instance, in the wake of the Coronavirus Disease (COVID-19) pandemic [67], many organizations had to cease any form of in-person gathering. This closure of spaces is a recommended non-pharmaceutical intervention (NPI) to limit spread of contagious diseases [68]. However, implementing these interventions can result in a variety of counter-productive side effects that impact both organizations and the individuals in them. Consider organizations like university campuses, such shutdowns can deprive smaller businesses that constitute the campus ((e.g., boarding, parking, dining, etc.) [69, 70]. Even for individuals, policy changes can lead to developmental challenges, such as learning loss for students when an organization moves all activities to an online format [71, 72]. Further, such reactive policies can leave members of an organization further isolated and deplete overall wellbeing; leading to anxiety and stress [73, 74]. Unfortunately, many such organizational decisions intended to regulate behavior are not derived from behavioral data. Through my work I illustrate that passive sensing frameworks can provide flexible opportunities to design data-driven policies for large organizations and work spaces.

2.3 Passive Sensing Frameworks for Digital Phenotyping of Work

Over the years, different studies in the field of HCI, Ubicomp, and CSCW have investigated the potential of passive sensing frameworks to understand worker behaviors and perceptions. This literature has enabled *in-situ* studies of workers without disrupting their naturalistic work routines. These technologies can overcome the challenges faced by survey instruments, to acquire appropriate representative data of an individual's dynamic context.

In this section. I discuss these prior works along the same multi-level framework used in the previous section (individuals, teams, organizations). The studies described below motivate my work towards more holistic understanding of worker experiences.

2.3.1 Sensing Individuals

As smartphones and wearables started becoming more popular, researchers started exploring ways to harness its sensors to describe human activity [75]. Early work in this space started with using inertial sensors embedded in these devices to identify basic daily activities like walking [76]. Eventually, this work was extended to tap into other sensing modalities (e.g., audio) to express mental states of users [77]. These frameworks could sense users unobtrusively and continuous because they relied on technologies that were naturally integrated into users' lifestyles. As a result, this type of passive sensing ignited a variety of *in situ* studies involving longitudinal behavioral analysis of individuals. For example, [78] demonstrate the potential of smartphone sensing to indicate the wellbeing and performance of college students [78]. Note, these frameworks are not limited to physical technologies that an individual interacts with. [79] have shown that online platforms, such as social media, can be used as passive sensors to infer mental wellbeing [79]. These studies indicate that passive sensing is capable of identifying a variety of human activity and mental health constructs. Therefore, it is natural to assume that these technologies reflect behavioral signals that could explain worker experiences.

In 2016, Mark *et al.* computationally analyzed the email interruption patterns of information workers and its relationship to performance and stress [80]. Smartwatches have also been studied to predict worker's cognitive load [81]. Several studies have taken a multimodal approach to estimating a worker's effectiveness. Large scale-sensor deployments have modeled the minutiae of daily behaviors to distinguish high and low performers at information work [82]. Other studies focused specifically at time at work, have been able to make task management suggestions to information workers based on passively sensed activity [83]. Complementary to this, studies of workers behaviors exclusively outside work have also indicated their performance [84]. While most of these works imply that a worker's activities (in and around their workday) can indicate their work outcomes, most of these studies do not distinguish these relationships from conventional indicators such as personality. Would single instance personality surveys still determine the same results without the need of extensive sensor deployment? Moreover, many of these studies do not clarify the activities that actually describe holistic workers. What does a high performer get right that can be recommend to others? Existing approaches to utilize passive sensing have not considered these domain driven questions.

In contrast to prior studies, I specifically aim to disentangle how, and to what extent, a worker's day-level activities are associated with performance independent of their personality. Instead of an intelligent model to distinguish workers, I want to demonstrate passive sensing as a worker-centered analytical lens to express positive daily activities.

2.3.2 Sensing Teams

Sensing an information worker's activities in isolation ignores the effects of social context on their work experience. Several studies have attempted to bridge this gap by using passive technologies for "socially aware computing" [85]. Using novel wearable sensors to detect face-to-face social interactions can explain aspects of job satisfaction [86, 87, 88]. Brown *et al.* modified the spatial layout of office spaces and studied the impact on interpersonal interactions through wearables [89]. Similar approaches can also explain the social clusters of between information workers [90]. Even digital behavioral traces, such as email interactions between information workers can reflect power relations [91]. Beyond interpersonal or dyadic interactions, other studies have been able to Other studies have been modeled individual worker behaviors to extrapolate norms of their teams [92, 93].

Although these studies analyze the social activities of workers with passive sensing, very few of these studies expand beyond dyadic interactions. Moreover, these studies have been insufficient in expressing the association between social interactions and work outcomes. I build on this literature by using passive sensing to study social behaviors within multiple groups of workers longitudinally. My work demonstrates how passive sensing of socially connected workers can shed light on their individual work experience.

2.3.3 Sensing Organizations

Certain aspects of an information worker's experience are the result of phenomena that permeate through the entire organization. Merely sensing individual workers, or even teams, is not sufficient to automatically determine normative behaviors, expectations or perceptions.

Prior work has suggested the potential of using social media as a passive sensor of organizational patters [94]. Information workers often use social media for a variety of purposes such as information seeking, knowledge discovery and management, expert finding, internal and external networking, and potential collaborations [95, 96]. In fact, many organizations even have internal social media platforms [97, 98]. Shami *et al.* proposed a tool, *Employee Social Pulse* — that analyzed streams of internal and external social media data to understand opinions and sentiment of employees [99]. Linguistic analysis of similar streams of employee social media has been successfully employed to gauge employee engagement [100, 101]. Although such social media has benefits, scholars argue that candid social media use within organizations can be affected by privacy-related concerns such as the breach of boundary regulations and employer surveillance [102, 103, 104]. Instead, I look towards public work-related social media to leverage self-contributed worker perceptions. By incorporating this kind of data into a passive sensing framework, I seek to aggregate large volumes of employee perspectives to describe affects of organizational culture on work outcomes.

Alternatively, organizational behaviors can be sensed or anticipated using different physical sensing technologies. For instance, studies have inferred campus mobility by accessing user devices with specialized data logging applications [105, 106, 107]. However, such approaches also fail to sufficiently represent the worker community because they require mass adoption and continuous maintenance of user devices. The lack of scale can be of crucial importance, especially when considering decisions to respond to organizational crises, such as an infectious outbreak. To mitigate these challenges, prior studies have illustrated the repurposing of already existing managed WiFi networks to model physical activity of the entire community [108, 109, 110]. My research furthers the potential of such passive sensing frameworks to flexibly determine behavior-driven organizational policies.

CHAPTER 3
DATA AND MATERIALS

To answer the research questions posed by this thesis, I have relied on a few datasets. The passive sensing frameworks that I have studied use some subset of these datasets and in some cases combine it with other data specific to that research goal. This chapter describes those datasets as a common reference.

3.1 The Tesserae Project

The *Tesserae* Project [111, 94] was a large scale, multi-university effort that extended prior efforts to use passive sensing to infer behaviors and mental states [112, 113, 79, 78]. The objective was to leverage the commercially available technologies and understand workplace performance longitudinally and in-the-wild.

A rolling enrollment from January 2018 through July 2018 led to a recruitment of 757 information workers. These participants were recruited from different field sites across the United States. Participants were either compensated by direct payments or through a set of weekly lotteries, based on the specific field site's requirements. Participants were requested to remain in the study for either up to a year or through April 2019. To infer participant activity and physiological context, various off-the-shelf technologies were provided to the participants; 1) Phone Agent—a smartphone application [78] to track phone usage (e.g., screen lock/unlock and GPS locations); 2) Wearable—a smartwatch (*Garmin Vivosmart 3*) to capture heart rate, stress, and physical activity, 3) Bluetooth beacons (*Gimbal*)—two static (to track their home and work location) and two portable devices (to carry on their person); 4) Social Media data [94]. Additionally, participants were requested to complete an initial set of psychometrically validated questionnaires related to demographics, job performance, personality, intelligence, affect, anxiety, alcohol and tobacco use, exercise, sleep, and stress, administrated via validated psychometric survey instruments. Participants were also required to complete a variety of surveys in regular intervals (daily, every other day, and weekly) throughout the course of the study to capture short-term fluctuations.

Demographics. A random sample of 154 participants (20% was "blinded" for external validation [111]. My findings are based on the remaining 603 "non-blinded" participants of which 253 reported they were female. On average a participant in the study was 34 years old (stdev. = 9.34). A majority of the participants possessed a college degree (52%), another significant proportion reported having a master's degrees or graduate level certification (43%), while few had doctoral degrees as well (4%). In terms of annual income, roughly the same proportion of participants earned between 50-100k USD (43%) and more than 100k USD (48%) per year.

Participant Privacy and Consent. Given the sensitive nature of this data, participant privacy was a key concern for the researchers. In addition to the informed consent form, researchers provided participants with a technical specification document that described the data sensed by each stream as well as methods to store and secure it. After reading this, the participants could specifically consent to each sensing stream they wished to provide data on. Participants could clarify their queries about the sensing streams through in-person discussions as well as e-mails. The data of enrolled workers were deidentified and stored in secured databases and servers which were physically located in one of the researcher institutions, and had limited access privileges. The study was approved by the relevant Institutional Review Boards.

3.1.1 Measures of Job Performance

Worker success was assessed along three different dimensions: task performance, organizational citizenship behavior, and counterproductive work behavior [32, 114]. Participants had to complete these surveys in an initial battery to evaluate their stable measures of performance. Additionally, they also had to complete versions of these surveys in more regular intervals to capture more time-sensitive aspects of performance.

- **Task performance:** This refers to behaviors typically rewarded by the management. It is characterized by a worker's proficiency at tasks that transform raw materials (objects, thoughts, or actions) into products or services [115, 114]. For example, in firefighting the task performance can be assessed on the basis of rescue operations, or, in software development it can be associated with bug fixes and feature deployment. While task performance focuses on a worker's accomplishment of prescribed duties, it does not account for their experiences during downtime (e.g., lateness), interpersonal interactions (e.g., collaboration) or destructive behaviors (e.g., plagiarism) [114]. The project used the Individual Task Proficiency (ITP) scale [116] and the In-Role Behavior (IRB) [117] scale.

- **Organizational Citizenship Behaviors:** These behaviors are not explicitly prescribed by management, but promote welfare within the organization [118, 119, 120, 121]. Such "extra-role" behaviors are voluntary and can either be *individually targeted* (e.g., aiding a peer) or *organizationally targeted* (e.g., volunteering in extra-professional activities) [122]. Contemporary notions of worker performance argue that at an aggregate, citizenship behaviors are as crucial as task performance in determining overall organizational outcomes [115]. Additionally, citizenship is one of the performance metrics that is related to factors like job satisfaction, which can predict organizational committment, turnover and absenteeism [123]. This construct was measured using the Organizational Citizenship behavior Checklist (OCB-C) [124] and a daily Citizenship Behavior scale [125].

- **Counterproductive Work Behaviors:** This refers to behaviors intended to jeopardize the organization or the individuals within it [126, 127]. Examples include stealing from a peer, insulting a colleague, purposefully doing tasks incorrectly. Recently, such behaviors have been considered to be a dimension of job performance because they can disrupt workflows of coworkers and the organizatin itself [126]. Sometimes this is considered as the most important metric when assessing overall workplace effectiveness [32]. This was captured using the Interpersonal and Organizational Deviance (IOD) scale [128] and a short-term instrument developed by Dalal *et al.*, which measure interpersonal deviance and organizational deviance. Throughout this document, I refer to these measures as “inter.deviance” and “org.deviance”.

Job performance can be measured by “self-report”, where a worker rates themselves. Alternatively, it can be measured by “other-report”, where a supervisor or peer rates the worker. While, the former can suffer from social-desirability effects, the latter gets inflated by *halo error* — rating based on overall impression bias as opposed to specific instrument categories [4, 5]. In practice, these approaches are not all that different. Meta-analyses of job performance instruments have shown that there is high convergent validity between the two methods [4, 5]. This project adopts validated self-report instruments to quantify each of our performance metrics.

3.1.2 Measures of Psychometric Characteristics

Beyond an employee’s task accomplishment, measuring their general wellness is important to infer their success in an organization. The wellness of participants was measured using both survey instruments and objective measurement of physiological changes.

- **Anxiety:** For an individual, anxiety reflects magnitude of subjective feelings like tension, apprehension, and nervousness. High anxiety is a marker of poorer worker wellbeing [129]. A single item instrument developed by Davey *et al.* was administered daily to compute fluctuations in anxiety[130].

- **Stress:** In the work context, stress can be viewed as the effect of external demands of one’s workplace [131]. The relationship between stress and job performance has been studied comprehensively in the past work [132]. This study used a daily single-item omnibus question to explore this phenomenon, *"Overall, how would you rate your current level of stress?"*. This instrument was internally validated within the program metrics of the overall project by robustly correlating it with other measures.

- **Arousal:** External stressors can lead to a "fight-or-flight" response in an individual. A common physiological response associated with both anxiety and stress is heart rate and heart rate variability. The participants' wearable devices employed an optical heart-rate (HR) sensor, combined with heart-rate variability (HRV) to compute their arousal score periodically throughout the day. This score was categorized as either "restful" or "stressful". The Garmin device measured the time a user spent in each state daily. According to *Firstbeat Technologies,* (the analytics behind Garmin's *HealthAPI* [133]) when an individual exhibits low HR and high but uniform HRV they are considered to be in a recovery state or at rest as the effect of the SNS on the body diminishes [134]. Typically this indicates relaxation, such as sitting or sleeping. On the contrary, when an individual's HR increases and their HRV drops below their baseline (rest) their SNS dominates, activating the body into a stress state. Experiments with Garmin reported that in free-living conditions, using HR to infer stress demonstrated low error (approximately 5% Mean Absolute Percentage Error) in estimating VO_2 max— maximal oxygen update, which is a key physiological indicator of stressful arousal [135]. External studies have also found that Garmin's HR based inference of VO_2 max were highly correlated ($r = 0.84$) with measurements from clinical instruments [136]. Garmin wrist worn devices have been used by researchers in the domain to provide physiological ground truth for modeling passively sensed behavioral signals [137, 138].

3.1.3 Passively Sensed Data

Unobtrusive, automatic and continuous activity data of participants was collected through different sensor streams, (i) smartphone, (ii) wearable, and (iii) Bluetooth beacons, (iii) so-

Table 3.1: Activity features derived from offline sensors; : Wearable, : Phone Agent, : Beacon
* : includes features aggregated by epochs, i.e, 24 hours, 12am - 6am, 6am - 12pm, 12pm - 6pm and 6pm - 12am

Category	Features	Stream
Activity Label	Still duration*, walking duration*, running duration*, unique activity count	
Movement	Steps count, steps goal, floors climbed, floors goal, distance covered	
Mobility	unique location count, total location count, inter-location distance	
Sleep	Sleep duration, sleep debt, time of wakeup, time of bedtime	 
Screen	Unlock Duration*, Unlock Count*	
Presence	Work session duration, desk session duration, desk session count, percentage time at work, percentage time at desk, 30–minute break count	
Colocation	Time of first and last interaction, number of interactions, number of unique participants, duration of interactions, percentage alone, percentage with at least one other/two others/three others	

cial media data. The application installed in the smartphone [78] measured screen activity (or device use), tracked GPS location, and provided activity labels [139]. The wrist-worn wearable estimated activity duration, step counts and was combined with the screen usage to yield sleep features. Lastly, the Bluetooth beacons were placed on the front door of the participant's residence and on their work desk. These beacons were observed by the phone agent [140] on the individual to infer the time they spent on their desk, when they came into work, and how frequent their time away from the desk was [16]. Table 3.1 summarizes the features derived from the physical sensors (for offline activity). These features are grounded in prior works of passive sensing [78, 141, 82, 79]

Language on social media was used to infer psycholinguistic attributes participant posts by using LIWC (Linguistic Inquiry and Word Count) [142] This lexicon has been used in prior work to study mental health and wellbeing through social media [143]. For my studies, I used 50 categories of LIWC that could segregated into the 9 different groups [143],

Table 3.2: Language features derived from social media

Category	Features
LIWC	*Affective attributes*: anger, anxiety, negative and positive affect, sadness, swear; *Cognitive attributes*: causation, inhibition, cognitive mechanics, discrepancies, negation, tentativeness; *Perception*: feel, hear, insight, see; *Interpersonal focus*: first person singular, second person plural, third person plural, indefinite pronoun; *Temporal references*: future tense, past tense, present tense; *Lexical density and awareness*: adverbs, verbs, article, exclusive, inclusive, preposition, quantifier; *Biological concerns*: bio, body, death, health, sexual; *Personal concerns*: achievement, home, money, religion; *Social concerns*: family, friends, humans, social
Sentiment	Positive score, negative score, neutral score
N-Grams	Top 500

affective attributes, *cognitive attributes*, *perception*, *interpersonal focus*, *temporal references*, *lexical density and awareness*, *biological concerns*, *personal concerns*, and *social concerns*. Additionally, posts were characterized with sentiment analysis (score for positive, negative and neutral label) [144]. Lastly, this data provided a large set of open vocabulary features, i.e., the usage of the top 500 n-grams [145] within the corpus of all posts in the study. These features were sparse because n-grams do not appear consistently on all posts but are still a mainstay in language-based predictions of mental wellbeing [146, 147, 148]. Table 3.2 summarizes these features extracted from social media.

3.2 CampusLife WiFi Logs

To understand the social dynamics and larger community-driven behaviors of workers, I leveraged a university campus' WiFi infrastructure. This network enables $40,000$ unique users every year to connect to the internet via $6,959$ different access points distributed between 204 different buildings. This helps conceive passive sensing frameworks that can utilize retroactively study behaviors and inform workplace decisions. Arguably, students within a university of higher-education are not working, as most of them are not expected to provide labor to the university. However, the overall behaviors of university students are not very different from information workers — their daily life involves information gathering, interpretation and creation (through classroom learning and projects). Moreover, the WiFi infrastructure also captures the behaviors of faculty, staff, researchers, and assistants. Therefore, to answer my research questions, I consider a university population as a proxy for a large organization using such data.

The IT management facility at Georgia Tech (GT) accumulates WiFi access point logs over time. This is common in most universities with managed WiFi infrastructure. The logs indicated the WiFi access point (AP) that a WiFi user's device was associated with. Thus, it can be used to infer dwelling locations of users across the entire campus. However, this approach is limited to indoor spaces where AP s are located and the scope of this localization is at the granularity of a room or suite [108, 149]). These logs do not contain any personally identifiable information and locations are also coded. I actively collaborated with IT management to define safety and security safeguards that allow us to obtain a deidentified version of these raw logs.

My research has leveraged this infrastructure for two different samples:

1. **Design Project Teams (Spring 2019)**: With approval from our Institutional Review Board, I obtained consent from 186 students to analyze deidentified versions of their WiFi association logs. They were also required to complete an entry survey. Participants were remunerated with a $5 gift-card for enrolling. 170 students were in the age range of 18-24 years, and 16 were of age 25 and above. Among these students, 59 reported female (32%). As per the official headcount 25% of the students within the CS major have been recorded as female. This sample's data was analyzed retroactively starting from January 1, 2019 for 14 weeks. Participants from this sample were used for developing the fundamental passive sensing framework and testing its reliability. More importantly, this population was used to investigate passive sensing to explain effectiveness in teams.

2. **All Visitors (Fall 2019 – 2020**: I was able to study a larger volume of logs to under-

stand organizational–, or in this case, community–patterns. This was made possible using a data-use agreement and a new ethics protocol that was approved by the Institutional Review Board (IRB). I analyzed logs from August 2019 through Dec 2020. Each day, approximately $33,000$ different people connected their devices to the WiFi network on campus. This population was entirely anonymous and this data was not combined with any other student-centric information. Overall, this data helped determine normative behaviors and drive organizational decisions.

Privacy. Participant privacy was a key concern for us. The two core streams of data, course outcomes and WiFi AP logs, are both de-identified and stored in secured databases and servers which were physically located in the researchers' institute and had limited access privileges. The study and safeguards were approved by the Institutional Review Board of the authors' institution.

3.2.1 Managed WiFi Network

Every AP installed on campus is mapped to a building ID and a room ID. The room ID indicates the room closest to the AP or the room that contains the AP (Table 3.3). Every entry in the log documents an SNMP (Simple Network Management Protocol) update in the network. This update is triggered when APs see a change, i.e., a device connects, or through an SNMP poll request to the AP that returns connected devices. Therefore, the log itself indicates that a device is in the vicinity of an AP, but without information of the client RSSI, this inference has a low spatial resolution. Moreover, the logs for a connected device are erratic because of variable connectivity settings in the device agent (e.g., the WiFi turns off when inactive). The irregularity in log updates leads to a low temporal resolution. The low resolution is what introduces "coarseness" to this data. Outside of the specific association timestamps—when an AP responds to an SNMP poll or a client switches APs—the connected device is invisible in the logs. For less than 5% of the APs, an AP shared a space label with another AP. This many-to-one mapping is typically in the

Table 3.3: Sample raw log

Field	Sample
Timestamp	Apr 1 00:10:51
Update Type	snmpupdate
Anon. User	2099
User Device	c4:7d:eb:0f:df:d5
AP ID	40:cd:14:b2:02:c0
AP Label	122S-209

case of large halls and auditoriums. I resolved such many-to-one associations by using APs as a proxy of the space they are associated with. Therefore, individuals connected to different APs in the same space will still be identified as collocated. Similarly, an individual could connect to the network with multiple devices. However, less than 1% logs show that a user is connected to multiple APs around the same time.

Coarse Localization. Like most universities, GT's managed WiFi network is not equipped with any Real-Time Location System (RTLS) [150, 151]. RTLS systems use Received Signal Strength Indicator (RSSI) values from multiple neighboring APs to provide high precise localization of individuals in terms of time and space. However, deploying such systems requires surveying the entire network. Additionally, precision localization raises more privacy concerns. These factors together make it challenging for universities to justify the deployment of RTLS, unlike small retail settings that can monetize RTLS insights directly (e.g., insights on footfall can be tied to improving revenue).

3.2.2 Inferring Mobility, Dwelling, and Collocation

The sample of design students included 2 sections (referred to as "1A" and "1B"), where the instructor provided lecture-by-lecture attendance information for the participants. I used the attendance records of these 46 students to inform the heuristics and test the reliability of the localization using these logs. By modeling logs accumulated in the 30 minutes before and after the lectures of sections 1A and 1B (Table 3.1), I was able to infer a user was moving when successive logs show connections to different AP s. By considering the 90^{th}

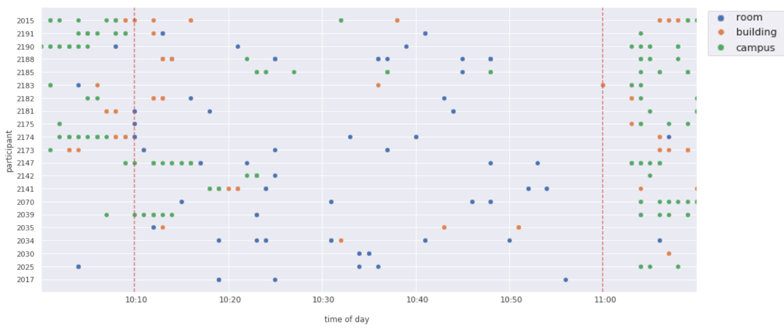


Figure 3.1: Markers represent log updates. The red vertical lines demarcate the lecture period. Consider Participant 2173. They associated with an AP outside the building, then logged an entry at an AP in the same building before logging an entry in the classroom, almost 8 minutes later.

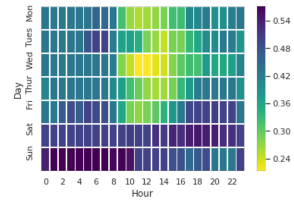


Figure 3.2: The median portion of time a user is disconnected from campus for a given hour for a day of the week

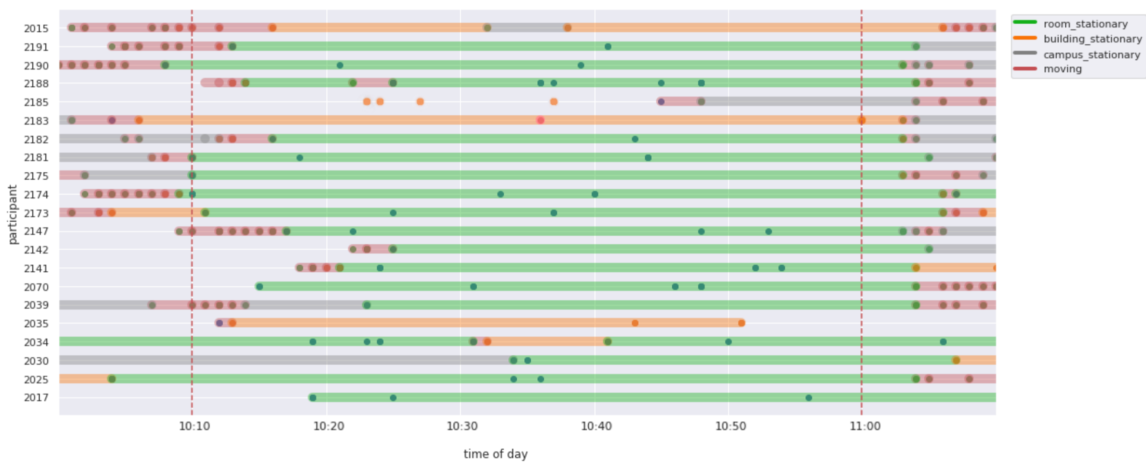


Figure 3.3: **Dwelling Segments**. The time periods between moving segments are interpolated as dwelling segments

quantile of the intervals between 2 different logs (233 seconds), I was able to determine if a device was roaming and by proxy, the user was moving.

The user was considered to be dwelling for any time segment when they are not mobile (Figure 3.3). However, a user could also be outside of the network, or disconnected from it, during two periods of movement. To account for this, I filtered out disconnection periods by identifying large intervals (greater than 76 minutes) between successive logs. Based on this heuristic, the disconnections were found to be more common weekends and before or after class times (Table 3.2).

In general, when two users were found to be dwelling near the same AP at the same time they were considered to be collocated for the overlapping period.

3.2.3 System Reliability

To quantify the reliability of this coarse localization technique, I evaluated the attendance of the 46 students in 1A and 1B for the 34 lectures that occurred in the sample data period. Even though every AP's coverage on campus might vary, when students collocate to work outside lecture times they typically gather in breakout rooms, empty classrooms, library spaces, or other similar indoor spaces. Hence, I considered presence in class a reasonable ground truth to evaluate the reliability of our proposed automated method.

Missing Data. On certain lecture days, I did not find any WiFi log entry for some students. The red stacks in Table 3.4 show the number of students per lecture with no log entries for section 1B. Comparing this to the attendance records showed that 93% of the times a student does not appear in the logs, they were actually recorded as present by the instructor. One possibility is that the student either had all their devices turned off or connected to a different network (e.g., cellular data, or the guest network). Every student in our sample had no WiFi log entries on at least one lecture they attended (the median was five lectures). Therefore, despite its pervasiveness, leveraging the managed network can still miss out on students who were actually present. For such occurrences, the automated method cannot ascertain presence or absence and therefore, I exclude these student records (for that lecture) from further analysis.

Accuracy. I considered a student to be in class if any time during class they were inferred to be dwelling at their respective lecture room's AP. I found 89% agreement between the instructor's record of who was present and our estimated record, a *precision* measurement. Also, the *false discovery rate* was 0.103. Therefore, WiFi logs rarely indicated a student was at a location when they were not physically present. The false positives were possibly the result of students failing to record their name on the attendance sign-up sheet, possibly because of showing up late to class. Alternatively, for every instance when the

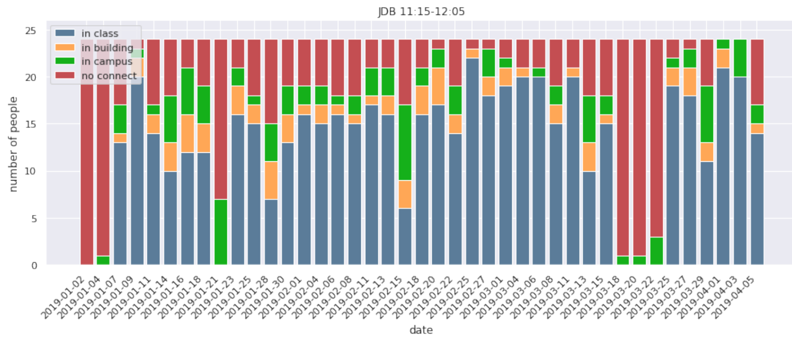


Figure 3.4: Each stack depicts where students were found to be connected — the lecture room's AP, another AP in the same building, to the campus network, or not connected at all.

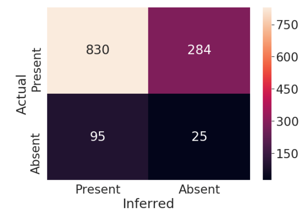


Figure 3.5: Actual vs inferred attendance — Precision: 0.89, Recall: 0.75

student was present, this method inferred them to be collocated 75% of the time—*recall*. For reference the *false negative rate* is 0.25. A false negative could occur when a student's device connects to a different AP on the network. Figure 3.3 denotes these as the orange segments. A device could also connect to an AP that is physically further away because the signal from their closest WiFi was attenuated [152].

To summarize, the *F1-score* of such a system can be interpreted as 0.81 (Table 3.5). It has high precision, but with a *specificity* of (0.74), it can erroneously mark students as absent when they were present. In the future, this can be addressed by deploying a broader set of APs for a given location.

CHAPTER 4
PASSIVE SENSING FRAMEWORKS TO EXPLAIN INDIVIDUAL WORKER OUTCOMES

A large part of an average information worker's routine actually involves thinking. Information workers cannot be gauged merely by production of commodities. Understanding what makes a worker effective needs a more holistic outlook that focuses on the workers motivations and expectations.

One common outcome that determines a worker's effectiveness is their performance at work. Theoretically, a worker's performance is considered a function of both their inherent personality and their daily activities [153, 154]. Yet, in practice, personnel management only incorporates personality-assessments to forecast job performance because traditional organizational research considers personality traits to be the most robust predictor of workplace functioning [155, 156]. However, personality assessments have limitations. First, the instruments used to measure personality rely on self-reports, which are vulnerable to feigning and self-presentation leading to subjective results [157]. Secondly, since changes in personality are only observed over long periods of time [158], its rigidity presents an inflexible view of a worker's job outcomes. In contrast, a worker's activities can be objectively measured and be flexible to changes. Therefore, by learning how actions determine workplace experiences, personnel management units would have the opportunity to recommend day-level activities to improve performance.

However, we have had little evidence indicating the relationship between daily activities and worker performance. Studying activities was once considered challenging because instruments like surveys, which rely on manual reporting, could not record the dynamic (moment-by-moment) *in situ* information nature of human activity. Passive sensing has changed this. Therefore, it is natural to ponder, how passive sensing can be extended to clarify activities of information workers.

RQ I: *How can passive sensing frameworks explain worker outcomes?*

In this chapter, I describe a study to empirically validate that personality and activity context are independently associated with job performance. I present the concept of an *organizational persona* — an intersection of two components; *activity*, which is sensitive to situational dimensions and demands; and *personality*, which is develops intrinsically. In my study, I demonstrate the use of off-the-shelf commercial technologies as a passive sensing framework combined with classical machine learning techniques to identify personas in a diverse organizational population.

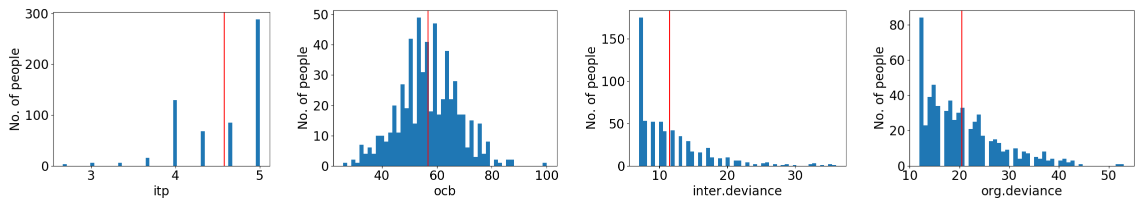
This chapter specifically documents:

- **Discovering Organizational Personas:** I describe an automated unsupervised clustering approach to discover meaningful organizational personas, which are composites of a trait-based *personality facet* and a dynamic, longitudinal *activity facet*.

- **Interpreting Organizational Personas:** I provide meaning to the four different personas I uncover by elaborating the various facets they are composed of.

- **The Role of Personas in Understanding Job Performance:** I examine how personas can reveal the association between an individual's activity context and their workplace functioning independent of pre-established relationships of personality to job performance.

Through my work, I show that passive sensing frameworkers can highlight organizational personas. These organizational personas provide a descriptive lens to interpret how daily activity data can complement personality to explain job performance.



(a) ITP (b) OCB (c) Inter.Deviance (d) Org.Deviance

Figure 4.1: The distribution of participants for different job performance variables with the red line indicating the mean

Table 4.1: Descriptive statistics of the job performance variables and personality traits

Measure	Mean	Std.Dev	Range	Scale
ITP	4.58	0.5	2.67-5	1-5
OCB	56.87	10.73	26-100	20-100
Inter Deviance	11.5	5.37	7-36	7-49
Org Deviance	20.48	7.56	12-53	12-84

(a) Job Performance

Measure	Mean	Std.Dev	Range	Scale
Neuroticism	2.46	0.78	1-4.92	1-5
Conscientiousness	3.89	0.66	1.92-5	1-5
Extraversion	3.44	0.68	1.67-5	1-5
Agreeableness	3.87	0.56	2.08-5	1-5
Openness	3.82	0.61	1.67-5	1-5

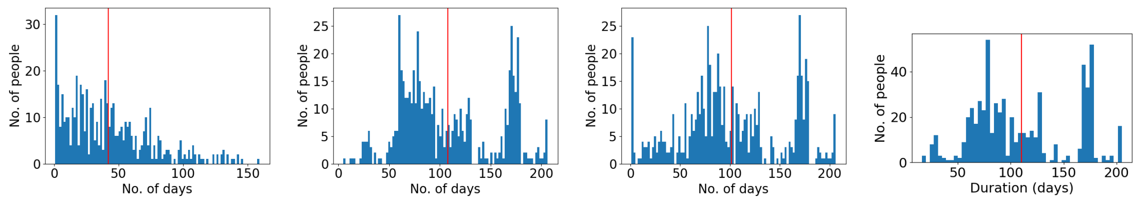
(b) Personality

4.1 Data

This study involves data from information workers who were recruited as a part of The Tesserae Project (section 3.1).

Self-Reported Data. The analyses presented in this chapter are only concerned with the personality metrics and job performance. Participants completed surveys to describe their personality based on the FFM(Table 4.1b). The personality measures were used to establish a baseline for performance in accordance with traditional approaches to evaluate information workers. I characterized job performance along three dimensions, task performance (ITP), citizenship behavior (OCB), and deviance. Participants were evaluated based on their response to different survey instruments (Table 5.7).

Passively-Sensed Data. To develop the passive sensing framework, I analyzed three different sensing streams — the smartphone, the wearable, and Bluetooth beacons. These streams described the daily activities of information workers at a daily level (Figure 4.2). For each participant, the feature values were summarized as the mean (of their study dura-



(a) Bluetooth Beacon (b) Wearable (c) Phone Agent (d) Study Duration

Figure 4.2: The distribution of participants by the number of days they provided data in each stream.

tion), in order to represent the sensed activity of their average day (Table 4.2). To describe the features, I have semantically segregated similar features into multiple sets :

- **Physical Activity**: The Phone Agent used the Google activity Recognition API [139] to measure the duration of different physical activities. Additionally, the wearable provided the distance traveled on foot.

- **Sleep**: The Garmin API provided information was supplemented with the phone unlock signals to infer an individual's sleep duration. Besides this, the dataset also computes sleep debt or deficit.

- **Mobility**: The number of locations a participant visited and the distance between these locations was calculated using the GPS coordinates logged by the phone agent. A DBSCAN clustering algorithm was used to identify significant locations across all the data-points.

- **Phone Use**: The Phone Agent recorded timestamps when the participant either locked or unlocked the screen. This was used to infer the number of unlocks as well as the duration of phone use (by measuring the time between an unlock and a lock).

- **Desk Activity**: The Gimbal API [140] in conjunction with the phone agent captured moments when particular beacons are within the range of an individual. Using the sightings of the static bluetooth beacon placed on a participant's desk, desk sessions and break sessions (of varying intervals) were inferred.

Table 4.2: 30 activity features representing the objective situations of the participant were derived and examined, this table shows a subset of them; : Wearable, : Phone Agent, : Beacon

Context	Representative Features	Stream
Physical	Distance On foot, Still Duration	
Sleep	Sleep Duration, Sleep Debt	
Mobility	Locations Visited	
Screen	Unlock Duration, Unlock Count	
Desk	Desk Session Duration, 30 Minute Break Count	

4.2 Discovering Organizational Personas

I operationalized an organizational persona as a combination of a *personality facet* and an *activity facet*. Activity is a mutable attribute, thus making it interesting to investigate in studies of performance and wellbeing. Contrary to personality, the malleable nature of activity presents the opportunity to manipulate it for experimental studies.

To obtain these facets, I employed a clustering method to identify mutually exclusive, homogeneous group of characteristics, which can be used to describe the participants. My approach centers on a specific fragment of context, *activity* — what someone does.

4.2.1 Feature Selection

At a formal level, the features I use express how an activity is performed (for e.g., duration, distance, count) and/or when it occurs (for e.g., throughout the day, from midnight to 6am, 6am to noon, etc.). The activity contexts shown in Table 4.2 can be extended on the basis of such qualities leading to a total of 30 features. Excessive features can add noise and latency to the modeling algorithm [159, 160]. Thus, to mitigate these effects I filtered out the most distinguishable features within my sample. Features like "outgoing calls" were rejected due to its sparse signal, i.e, event-contingent activities that occurred less than once a day were dropped. This was followed by a step-wise *Variance Inflation Factor* method [161] that eliminated features with high multicollinearity, for e.g., the phone unlocks from

Table 4.3: Standard coefficients of activity features when using multiple different regression models for each activity (an example of the variable-centered approach). These are useful for describing the variables but not the individual.
$M\ :\ performance\_metric\ \sim\ personality\_traits + activity_i$ ('-':p¡1, '.':p¡0.1, '*':p¡0.05, '**':p¡0.01, '***':p¡0.001)

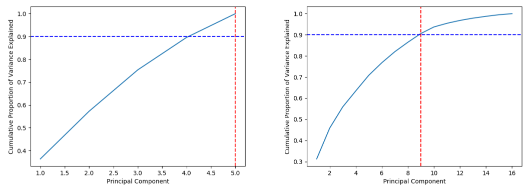
Activity Feature	ITP	OCB	Inter Deviance	Org Deviance
Locations Visited (12am-6am)	-	-	3.61×10^{-1}*	6.58×10^{-1}**
Locations Visited (12pm-6pm)	-	-	2.77×10^{-1}*	5.25×10^{-1}**
Distance-On foot	-1.85×10^{-5}*	-3.82×10^{-4}*	-	
Unlock Duration (12am-6am)	-	-	-2.31×10^{-4}*	-
Unlock Duration (6am-12pm)	-	-	-5.39×10^{-4}**	-
Unlock Duration (12pm-6pm)	-	-	-5.74×10^{-4}**	-
Unlock Count (12am-6am)	-	-	-	1.12×10^{-1}* -
Unlock Count (6am-12pm)	-	-	-	3.78×10^{-2}* -
Unlock Count (12pm-6pm)	-	-	-	5.49×10^{-2}* -
Desk Session Duration	-	-8.66×10^{-4}*	-	-
30 Minute Break Count	-	3.20×10^{-1}*	-	-

6pm to 12am can be explained by other features. For this step, a VIF threshold of 10 was used and the topmost feature was removed successively. Carefully curating features that depict activities followed by a few computational heuristics helped minimize the activity information to a set of 16 features.

4.2.2 Person-centered Approach

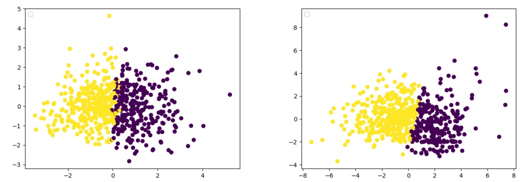
The selected features were fed into clustering algorithms to characterize the dominant patterns of personality traits and day-level activities. The use of classical clustering methods is motivated by a methodological perspective known as the *person-centered* approach [162]. Not only is "clustering" a popular method in organizational research [162, 163, 164], the person-centered approach views the individual as an "integrated totality" [164]. This is an alternative to *variable-centered* approaches that study human-centered data since those methods often treat individuals as a collection of isolated features [162].

Compared to variable-centered methods, person-centered methods can simplify the main effects and interactions of a large set of features through cluster representations. For example, we can consider taking a variable-centered procedure to unpack the role of per-



(a) Personality (b) Activity

Figure 4.3: 5 PCs for personality and 9 for activity are needed to explain 90% variance in their respective feature spaces

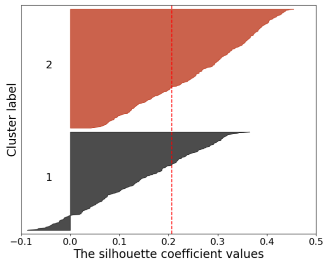


(a) Personality (b) Activity

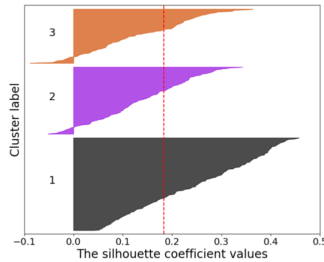
Figure 4.4: A representation of k-means clustering of the data along the first 2 PCs — explaining approximately 60% variance in the personality and 40% variance in activity.

sonality and activity by building multiple linear regression models. These models test the relationship between activity features and job performance metrics after controlling for personality traits. Table 4.3 shows which of the 16 shortlisted features exhibit significant associations with performance through the lens of a variable-centered method.

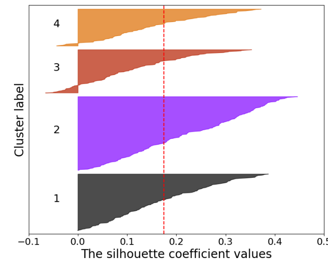
However, given a large number of separate models corresponding to each of the features, this framework is susceptible to false positives [162]. Theoretically, a variable-centered approach should account for multiway interactions between the features, but this requires including multiple interaction terms (multiple two-way terms, multiple three-way, and eventually n-way terms) [162]. Therefore, in studies that describe individuals along multiple-dimensions, regression models can often get bloated and challenge interpretability. On the contrary, taking all the activity features as a whole is more conceptually grounded and domain-driven [162] way to capture the interplay of dimensions. A person-centered method is a minimalist technique to meaningfully understand the nature of the individuals in a sample by capturing the interactions within the features concurrently. [1].



(a) $k = 2$



(b) $k = 3$



(c) $k = 4$

Figure 4.5: Silhouette plots for different k in the personality space. The red vertical line represents the average silhouette score

4.2.3 Clustering Procedure

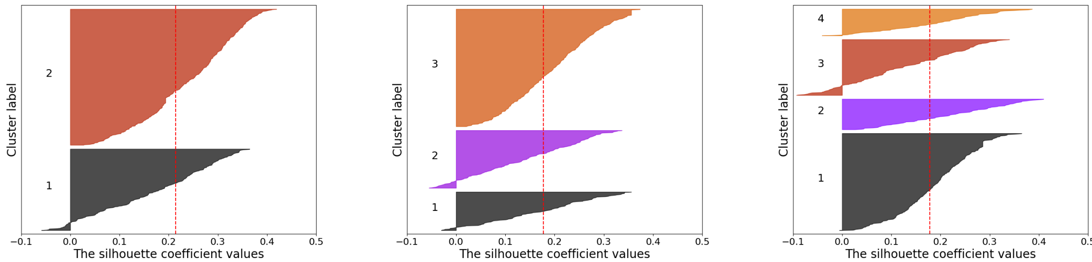
Before proceeding with clustering, I transformed the selected features for the best possible algorithmic results. To make the comparison across dimensions more equitable, I standardized the features into Z-scores [166] – the number of standard deviations a participant differs from the mean. This was followed by a *Principal Component Analysis* (PCA) to overcome the loss of precision in distance measures of high-dimensional spaces [167]. This dimensionality reduction technique transforms data into orthogonal dimensions. Based on a 90% threshold for explanatory variance I retained all 5 principal components in the personality facets the first 9 for the activity facet (Fig Figure 4.3b and Fig Figure 4.4).

Given my unsupervised approach, I first investigated which clustering method would be most suitable for the data. I tested these methods on the different feature spaces (personality and activity). The K-Means method was chosen and then applied to both feature spaces based on the best average silhouette score [168] (Table 4.4).

For K-Means, to determine the number of clusters I used the Silhouette Method [168]. Figure 4.5 and Figure 4.6 visualize the silhouettes based on how closely each point in a cluster is matched to the cluster. I found that the highest average silhouette score in both cases is at $k = 2$.

To summarise, I identified two distinct clusters in the personality feature space and

[1]Note: This process is not meant to supersede regressions, it is simply an alternate lens to examine individuals [165]



(a) $k = 2$ (b) $k = 3$ (c) $k = 4$

Figure 4.6: Silhouette plots for different k in the activity space. The red vertical line represents the average silhouette score

Table 4.4: Silhouette scores per clustering algorithm; k represents the number of clusters

	Activity score	k	Personality score	k
K-Means	0.214	2	0.207	2
Hierarchical	0.205	2	0.195	2
Affinity Propagation	0.074	47	0.153	46
DBSCAN	-0.289	2	-0.241	2

Table 4.5: Number of participants in each persona

	P_1	P_2
C_1	120	106
C_2	210	167

two dominant groups in the activity feature space for the participants. Each feature space corresponds to a *facet* in the organizational persona. For simplicity, I labeled the personality facets as $P_1$and P_2; similarly, the activity facets were labeled as $C_1$and C_2. Table 4.5 shows the distribution of the participants across the personas.

4.3 Interpreting Organizational Personas

Fundamentally, I envision an organizational persona as a construct that encapsulates the most discernible aspects of inherent traits (personality) of dynamic states (day-level activities) shared by a population of workers.

4.3.1 Meaningful Representation of Clusters

The high dimensionality of my feature spaces can make interpretation of clusters challenging [169]. I performed an ANOVA comparing the cluster means for each feature. If the F-statistic value (ratio of between-group variance and within-group variance) was sig-

Table 4.6: Results of ANOVA, comparing each cluster along different features; only features with F statistics significantly greater than 1 were retained; : Wearable, : Phone Agent, : Beacon

Feature Name	Source	p-value	F-stat
Explanatory Features			
Locations Visited (6am-12pm)	Phone	2.55×10^{-97}	645.92
Locations Visited (12am-6am)	Phone	2.17×10^{-74}	445.35
Locations Visited (12pm-6pm)	Phone	1.33×10^{-71}	423.28
Unlock Duration (12am-6am)	Phone	1.20×10^{-51}	278.44
Unlock Duration (6am-12pm)	Phone	1.60×10^{-43}	225.55
Unlock Duration (12pm-6pm)	Phone	2.23×10^{-41}	212.15
Unlock Count (12am-6am)	Phone	4.07×10^{-37}	186.15
Unlock Count (6am-12pm)	Phone	3.27×10^{-25}	117.96
Unlock Count (12pm-6pm)	Phone	9.44×10^{-22}	99.30
Desk Session Count	Beacon	2.38×10^{-16}	71.23
Sleep Duration	Phone, Wearable	5.10×10^{-14}	59.50
Still Duration (6am-12pm)	Phone	1.59×10^{-11}	47.21
30 Minute Break Count	Beacon	2.90×10^{-8}	31.60
Desk Session Duration	Beacon	3.85×10^{-8}	31.02
Pruned Features			
Sleep Debt	Phone, Wearable	1.03×10^{-1}	2.66
Distance On-foot	Wearable	5.60×10^{-1}	0.34

nificantly greater than 1, that feature was considered to sufficiently discriminate between clusters and thus non-trivial in describing a cluster [170]. For the personality clusters, all the features were retained, i.e. every trait was a good discriminator among the clusters. For the contextual clusters, however, two features were pruned out and the remaining 14 features were used to interpret these clusters (Table 4.6).

4.3.2 Personality Facets

A descriptive summary of different personality composites in our sample replicates constructs within established person-centered typologies, such as the ARC taxonomy [171]:

- P_1— *High Conscientiousness, openness, extraversion and agreeableness, Low Neuroticism*:

  This group scored high on all the personality factors except neuroticism. This cluster resembled the "Resilient" personality type based on the ARC taxonomy [171]. This

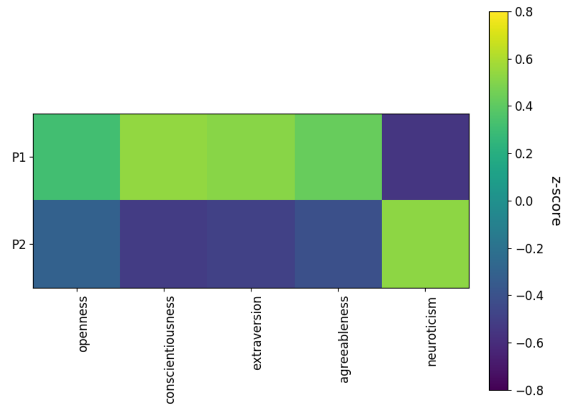


Figure 4.7: The Z-Scores of the personality clusters help distinguish the relative difference between the clusters features

Table 4.8: Absolute value of personality features across clusters
▲: values higher than mean, ▼: values lower than mean

	P_1	P_2	Mean
Openness	3.92 ▲	3.69 ▼	3.82
Conscientiousness	4.20 ▲	3.50 ▼	3.89
Extraversion	3.73 ▲	3.09 ▼	3.44
Agreeableness	4.09 ▲	3.61 ▼	3.87
Neuroticism	2.01 ▼	3.00 ▲	2.46
Participant Count	330	273	603

type is known to be most robust to psychological adaptability. In their approach, Gerlach *et al.* demonstrated that this composition of traits is often considered desirable, making it a role-model class [172].

- P_2— *Low Conscientiousness, openness, extraversion and agreeableness, High Neuroticism*:

  These individuals are generally the polar opposite of P_1in relation to conscientiousness and neuroticism. Their conscientiousness was 16.67% less than $P_1$and neuroticism was 49.2% higher than P_1. In terms of the ARC taxonomy, this group is similar to the "Overcontrolled" type that tends to be relatively antisocial [171].

Table 4.7 shows how each cluster varies along different traits with Table 4.8 depicting their absolute values.

4.3.3 Activity Facets

Unlike the clustering of information workers based on personality traits, the activity-based clustering has not been explored. This section describes the clusters in terms of the confluence of action-strategies participants employed to adapt to daily contexts. The dominant activity patterns in the passively-sensed data are described as follows:

- C_1— *High Mobility, Interruptive Phone Use, High Desk Dwelling, Low Sleep*:
  Individuals exhibiting this activity facet are relatively the most mobile. We see this pattern practically throughout the day (across all three 6-hour windows). In the morning from 6am to 12pm, members of this group visit about 7 distinct locations. It is worth pointing out that this does not necessarily imply a greater amount of physical activity since the average time they remain still is only 12 minutes less than the average. Similarly, this group was almost indistinguishable in terms of distance covered (note the poor $F-$statistic on Table 4.6). This gives us reason to believe that commute might be one of the cardinal activities of their day.
  Another distinct aspect of this group is interruption-heavy phone use. Since members of C_1exhibit lower duration of contiguous phone use and complement this with a high number of device unlocks it could indicate a greater disposition to interruptions. Despite these digital interruptions, their duration at desk exceeds the average. They also exhibit fewer instances of leaving their desk (*break count*).
  On average they sleep 20 minutes less than the mean. This could either be because of more commute (visit about 1 location more than average at night) or because of interruptions (unlock the phone 5 more times than average).
  Considering the concurrence of such action markers, C_1could represent workers who travel frequently and mostly work from the desk where they collaborate or communicate often, but remotely.

- C_2— *Low Mobility, Batched Phone Use, Transient Desk Dwelling, High Sleep*:
  Members of C_2 log far fewer distinct locations throughout the day. The amount of time they are still is slightly higher than average, but their physical movement is not significantly lower (Table 4.6).
  Interestingly, in contrast with to C_1, these individuals demonstrate contiguous phone use or "batching" [80]. They unlocked the phone fewer times than usual, but their duration of use was longer. For example, from 12am to 6am, they unlocked their

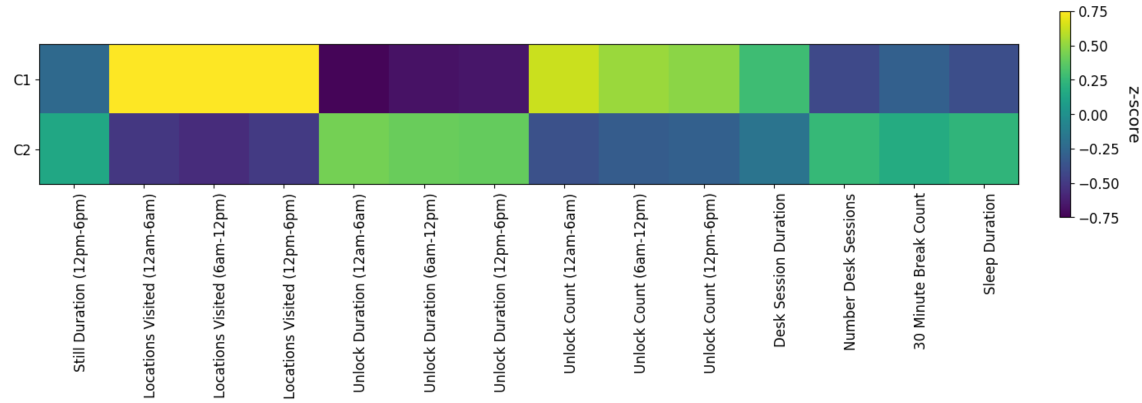


Figure 4.8: Heatmap showing the Z-Scores of contextual clusters

phones around 6 times less than C_3, but they used the phone for 35 minutes longer. Another divergent aspect of this cluster is their desk activity is an inverse of the previous group. They log far more desk sessions, but the average duration is shorter and the number of breaks is higher. This ties in with their phone use patterns as well. In addition to these qualities, this set of individuals end up sleeping longer as well. Even though they use their phone longer, the fewer interruptions could indicate better resting due to lack of interference.

These workers spend most of their time in one place but not necessarily at their desk. This implies more frequent physical collaborations within a site, and digital tasks being coordinated into chunks.

Figure 4.8 shows how each of these groups differs for different contextual factors. The absolute values for the activity attributes of these clusters can be referred to in Table 4.9.

4.3.4 Combining Facets to Construct Personas

Each personality facet combines with each activity one to give rise to a persona. With 2 facets of each kind, this approach renders 4 personas, referenced subsequently in the form of P_iC_j. I performed a χ^2 test to measure the association between the two facets and found no significant association ($p = 0.59$) between them. Therefore, individuals of

Table 4.9: Absolute value of activity features across clusters; ▲: above mean, ▼: below mean

	C_1	C_2	Mean
Physical Activity			
Still Duration (12pm-6pm), *minutes*	433.23 ▼	453.43 ▲	445.43
Mobility			
Locations Visited (12am-6am)	3.20 ▲	1.37 ▼	2.11
Locations Visited (6am-12pm)	7.36 ▲	2.43 ▼	4.44
Locations Visited (12pm-6pm)	3.87 ▲	1.58 ▼	2.51
Distance-On foot, *meters*	4836.65 ▲	4713.79 ▼	4760.69
Phone use			
Unlock Duration (12am-6am), *seconds*	568.53 ▼	2671.83 ▲	1846.57
Unlock Duration (6am-12pm), *seconds*	293.59 ▼	1418.34 ▲	978.56
Unlock Duration (12pm-6pm), *seconds*	332.28 ▼	1361.40 ▲	957.61
Unlock Count (12am-6am)	12.06 ▲	6.37 ▼	8.59
Unlock Count (6am-12pm)	40.53 ▲	24.08 ▼	30.51
Unlock Count (12pm-6pm)	21.19 ▲	12.79 ▼	16.08
Phone use			
Desk Session Duration, *seconds*	1379.38 ▲	827.86 ▼	1040.65
Number Desk Sessions	15.02 ▼	31.37 ▲	25.01
30 Minute Break Count	3.70 ▼	5.10 ▲	4.55
Sleep			
Sleep Duration, *hours*	6.84 ▼	7.36 ▲	7.16
Sleep Debt, *hours*	1.38 ▲	1.30 ▼	1.33
Participant Count	226	377	603

the same personality facet are equally likely to exhibit qualities of any of the two activity facets; and vice-versa. This is an important result to establish independence between our two constructs before comparing its effect on job performance.

Without the distinction mentioned above, it would be a struggle to elucidate the effect of personality and activity patterns on job functioning independent of the other factor. This two-faceted approach provides the flexibility to test the isolated effect of each factor as well as the interaction effect of each cluster.

4.4 Using Organizational Personas to Analyze Job Performance

This section will detail how to evaluate organizational personas in a $m \times n$ design where m is the number of personality facets and n represents the different types of activity configurations.

4.4.1 Significant Main Effect

Analyses

The two personality facets along with the two activity facets inform a 2×2 factorial design for ANOVA. This analysis will help reveal:

- If personality type effects individual job performance irrespective of their daily activity pattern — *Do personality clusters from our person-centered approach replicate relationships in the literature?*
- If day-level activity pattern effects individual job performance independent of their personality type — *Do regular dynamic-activity patterns add new information over and above personality type?*
- If personality and daily activity have a combined effect different from the sum of their whole in determining job performance — *Do the two independent factors (personality and activity context) interact significantly?*

.

I used a two-way non-parametric ANOVA test to compute the main effects of a worker's personality facet and activity facet with respect to the 4 job performance measures we collected from the initial enrollment survey (ITP, OCB, interpersonal deviance and organizational deviance). Specifically, I employed the *Aligned-Rank Transform* test [173, 174] — a robust measure that accounts for non-normality and can accommodate multiple categorical variables as independent variables (personality facet and activity facet).

Findings

The results of the previously described tests are recorded in Table 4.10. As prior work in organizational studies has posited, the personality facet of a worker is, in fact, potent in distinguishing job performance — across all 4 metrics, task performance (ITP),

Table 4.10: Significance of main effect and interaction effect of different facets of persona on measures of job performance
('-':p¡1, '.':p¡0.1, '*': p¡0.05, '**':p¡0.01, '***':p¡0.001)

	ITP	OCB	Inter.Deviance	Org.Deviance
Personality	8.03×10^{-17}***	0.028*	7.79×10^{-6}***	3.03×10^{-12}***
Activity	0.031*	0.026*	-	0.012*
Personality:Activity	-	-	-	-

citizenship (OCB), interpersonal deviance (Inter.Deviance) and organizational deviance (Org.Deviance) [175, 34, 37].

Additionally, it is also evident that an individual's activity facet does hold significance in explaining ITP, OCB and Org.Deviance (Table 4.10). This implies that even on holding out the effect of a worker's composite personality, their day-level activity patterns do describe their performance.

It is also important to note that our analyses found no interaction effect between the two patterns that represent a persona (Table 4.10). This result is key in our interpretation of the effect of each facet, as the lack of interaction indicates their effects are not confounded.

4.4.2 Effect Size

Analyses

I have already illustrated that passively sensed activity data can provide new information to explain workplace experiences. This section quantifies that effect, i.e., how much new information can be accounted for by situational activities. Considering the mixed-factorial design of the ANOVA, I computed the η^2 and the partial-η^2 for the personality and activity facet [176, 177, 178].

Findings

The ANOVA depicted that a worker's activity facet explains task performance, citizenship behavior, and organizational deviance. Table 4.11 records the η^2 and the partial-η^2 of both facets. The partial-η^2 approximates the effect of a facet if we had a single independent

Table 4.11: Effect size of the two different organizational persona facets in explaining different performance metrics.
η^2 indicates the effect of each factor when the individual is described in terms of both personality and activity.

	ITP		OCB		Inter.Deviance		Org.Deviance	
	η^2	partial η^2	η^2	partial η^2	η^2	partial η^2	η^2	partial η^2
Personality	0.954	0.954	0.434	0.855	0.982	0.987	0.855	0.972
Activity	0.063	0.784	0.492	0.870	0.005	0.253	0.120	0.828

variable (IV) in our ANOVA, without controlling for other effects. Since we use 2 IVs (personality and daily activity), the η^2 reflects the individual effect of each factor when both accounting for both facets of a persona.

In terms of task performance (ITP), both personality and activity have a large effect — partial-$\eta^2 = 0.92$ for personality and partial-$\eta^2 = 0.78$ for activity. However, the η^2 value infers that personality is alone responsible for explaining most of the variance in task performance. Once those effects are held out, the activity facet still retains a significant but small effect. Though the incremental delta of information the activity facet provides over a worker's personality is minimal, it still helps improve our understanding of task performance.

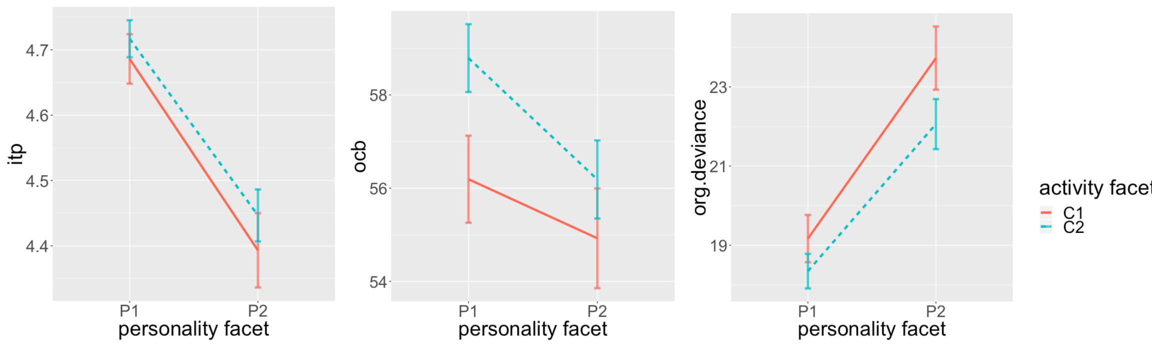
For citizenship (OCB), the individual effects of personality and activity are very similar, partial-$\eta^2 = 0.85$ and partial-$\eta^2 = 0.87$ for activity respectively. Evaluating the η^2 value elicits a similar result. These values suggest that considering the activity facet actually explains close to 50% ($\eta^2 = 0.49$) of citizenship behavior.

The tests on organizational deviance again show a large effect of personality, but there is still a considerable effect of an individual's activity facet. On noting the effects of both the factors, personality accounts for a majority of the variance ($\eta^2 = 0.85$) but activity does help complement this effect by explaining a significant segment of the remaining variance ($\eta^2 = 0.12$). The incremental variance explained by the daily dynamic activities represents the proportion of performance that can be varied with changes in activity variation.

4.4.3 Interpretation of Effects

I have shown that passive sensing enables a person-centered approach that provide a descriptive lens to understand these effects.

Fig Figure 4.9 illustrates how each of the four personas varies in terms of different job performance metrics. While the persona construct retains the dominating relationship between personality and performance, it uncovers new relationships of activities and worker functioning. Among the two facets, the activity composite depicted by C_2 (low mobility, batched phone use, and high sleep) appears to have more aspirational qualities. Personas with this facet had higher task performance and citizenship along with lower deviance. In comparison to the C_1, personas with this facet visited fewer locations in the day. Given the duration of stillness of this facet is comparable to the average, the greater number of locations could represent commute. Prior research shows that greater amount of commute and higher variability can lead to stress, strain and also an inability to perform non-work responsibilities [179, 180]. In C_1's case, the complex commute is also indicative of poor satisfaction [181] which in turn is related to poor citizenship behaviors at the workplace [123]. The patterns of phone-use by individuals in C_1 indicate interruptions. These could be self-initiated (distractions) or be external notifications (intrusion). Distractions can encourage procrastination and intrusions lead to disrupting task-flow that gives individuals the perception that their performance is lagging behind [182, 183]. In contrast C_2 have more compartmentalized use of their phones that could support to better task organization. While phone-use could cause interference, breaks from desk represent greater social interaction and collaboration that generally tempers deviant behaviors [184] and boosts citizenship [185]. These findings show how individuals sharing similar personality profiles could still differ in performance when they are involved in different daily activities.



(a) ITP (b) OCB (c) Org.Deviance

Figure 4.9: Main effect plots showing the effect of the 2 IVs (personality facet and activity facet) on job performance. Personas exhibiting P_1 personality score better on all metrics. Personas demonstrating C_2 activity patterns rate better on all metrics.

4.5 Summary

By leveraging passive sensing I demonstrate a novel analytical lens, known as *organizational personas*, to inspect job performance of information workers from a unique descriptive perspective. I demonstrated how this composite construct furnishes evidence that a worker's activity context can explain workplace performance beyond the static effects of implied by their intrinsic personality. I described how to identify and meaningfully describe these personas using standard clustering methods. Furthermore, I demonstrate that these personas help us understand the extent to which a variation in mutable activities can explain performance. I believe this work encourages harnessing pervasive technology embedded in a worker's everyday to gather activity information and in turn improve the understanding of personnel performance as well as inform subsequent efforts to alter it.

CHAPTER 5

PASSIVE SENSING FRAMEWORKS TO EXPLAIN GROUP DYNAMICS AT A WORKPLACE

Workers and their functioning at work is often judged in isolated silos. In fact, the studies of personality also reinforce this singular focus on the individual. However, on ground, information workers often work collaboratively and their experiences are the result of social interactions with their peers [12, 186]. Even when information workers are not directly working together, their presence in the same space or awareness of each other leads to mutual exchange of information [44]. Apart from enabling serendipitous interactions, workers can feel more motivated, secure and also up to date when other workers are nearby. The behavior of an information worker is often oriented towards the behaviors they observe. As a result, the success of a worker can be related to interpersonal dynamics at work.

Measurement of group dynamics presents new challenges which would not be encountered if one were to simply evaluate an individual worker. While groups of workers have been ethnographically observed it is difficult to expand the observations across multiple groups. Alternatively, I/O psychology researchers have considered comparing individual perspectives with social norms to indicate P–O Fit [48]. Yet these measures of social fit tend to be rigid and do not reflect dynamic changes or opportunities of improvement [56].

In the previous chapter, I already demonstrated a passive sensing framework to clarify daily activities of high performing information workers. As a natural progression, I now seek to answer how passive sensing can clarify experiences within a socially connected set of information workers.

RQ II: *How can passive sensing frameworks explain social dynamics at a workplace?*

In this chapter, I discuss two studies to answer this research question. Both studies rely on sensing group behavior with sensors embedded in the infrastructure of an information worker's environment. First, I repurposed managed WiFi networks to analyze collocation patterns of socially related workers to provide empirical evidence for social correlates of performance. Second, I leveraged static Bluetooth beacons to model work routines of socially related workers to provide novel measures of social fit.

These studies highlight the potential for passive sensing frameworks to: (i) empirically validate theoretical phenomena related to performance within group–settings, (ii) conceive new measures of team cohesion related to the performance and wellbeing of workers.

5.1 Leveraging WiFi Network Logs to Infer Collocation of Teams and its Relationship with Performance

At work, it has been observed that being collocated in the same space provides common artifacts for reference and helps collaborators coordinate their effort [44]. Additionally, collocation provides the opportunity for synchronous interactions through multiple channels — voice, expressions, gestures and body posture — and for impromptu interactions that strengthen social ties. However, empirical assessments of these behaviors with traditional surveys is obtrusive and does not scale over time or space.

The passive sensing community has introduced many automated and unobtrusive sensing methods to capture collocated social interactions [85, 187, 106, 188, 189]. However, most approaches that require specialized client-side applications [106, 187, 190] require collective adoption from multiple socially related workers. Even infrastructure-based techniques do rely on client side components to process the data [106, 191, 189] and equipping the surroundings with new embedded sensors. Harnessing worker's devices to infer social interactions in this way, while tempting, is challenged by privacy concerns, power consumption, and maintenance issues. Alternatively, embedding new sensors in the environment is limited by the expense of covering the entire campus. Moreover, none of these approaches can retrospectively study social behaviors over key-events and periods of interest. Together these factors challenge the scalability of such methods because they provide a sparse representation of the community.

In contrast, many campuses maintain a managed WiFi access-point (AP) network that provides device association logs which can be repurposed to infer locations of users [192] and subsequently model individual behaviors [109, 108]. Albeit a coarse descriptor of location — with low spatio-temporal resolution — these WiFi association logs can describe collocation of individuals. *Positing that these collocation behaviors present avenues for* social interactions, in this study, I examine their relationship to the performance of students in project groups.

Specifically, I pursue the following research goal:

To what extent is WiFi based coarse collocation associated with group members' academic performance?

Despite its low spatio-temporal resolution, I explore if unobtrusively inferred collocation of project group members is related to performance in the project. When individuals with a common intent gather in a space, it can describe their relationship to each other. In this work, I harnessed this aspect of human interactivity known as *spatiality* [44], by studying collocation behaviors of a set of university students that are known to share situated experiences on-campus. I validated my approach by examining, using statistical modeling approaches, if a student's collocation patterns were associated with an established outcome of social interactions—performance in teams [193, 194, 43, 42, 195].

5.1.1 Participants and Data

Computer Science driven Design Teams

As a proxy to information workers, I recruited participants enrolled in an undergraduate design course for CS students. The course is offered every semester and is a two-semester sequence. Students in this course were expected to work with a team of four to six students over two semesters (Part 1 and Part 2) on a single design project. In Spring 2019, this

Table 5.1: Participants in the study with complete data

Section	Part 1	Part 2
A	22	21
B	24	27
C	18	31
D	20	12
E	-	11
Total	84	102

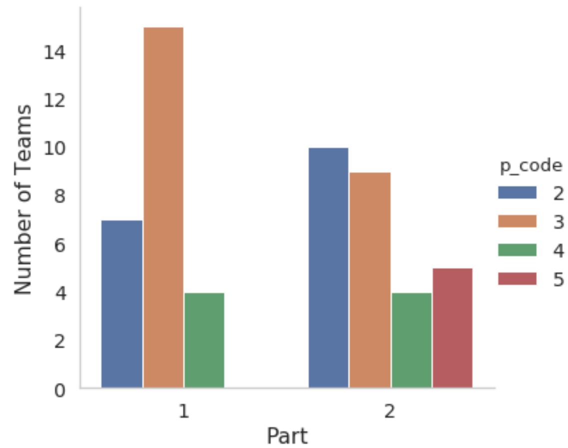


Figure 5.1: **Distribution of group sizes.** Among the students recruited, at least one other member of their group must consent.

course had four sections for Part 1 and five sections for Part 2.

Each section had an enrollment of about 40 students. In terms of course structure, Part 1 involved both lectures as well as project milestones. In contrast, Part 2 had fewer lectures and expected students to allocate scheduled class-times for project-related efforts. Students in both parts were expected to collaborate on project work outside scheduled lectures. It is not generally known how often student teams met outside of class, nor is it known how much those collocations impacted performance.

Upon enrollment, participants provided consent for the researchers to access their anonymized WiFi AP log data as well as their course data after completion of the semester. During enrollment, participants also completed an entry survey where they reported their group ID along with describing when, where, and how often they interacted with their group members face-to-face for class purposes. 186 students enrolled into this study.

Course Data

The course instructors provided course-related data for the consenting students along with course lecture times (Table 5.1). Among these students, 23 students did not have any other

Table 5.2: Peer-Evaluation Scales (1-5); Psychological Safety (1-7)

Construct	Mean	Med	Std
Member Effectiveness	4.36	4.45	0.51
Team Satisfaction	4.44	5.00	0.76
Conflict (Task)	1.64	1.67	0.62
Conflict (Relation)	1.26	1.00	0.51
Conflict (Process)	1.41	1.00	0.59
Psychological Safety	6.12	6.29	0.80

member from their group in our study and thus were dropped from this analysis. These remaining 163 students were in 54 separate groups (Table 5.1).

Final Score. This is a numerical score between 0 and 100 that informs the eventual letter grade based on the instructor's grading scheme. This final score is dominated by the project outcomes but students are assessed individually. These variations are introduced by participation as well as the instructor's subjective assessment of peer evaluation. Among the recruited group members, the range of scores between members could be as large as 6.5 points. This final score represents the ground truth for a student's academic performance.

Peer Evaluation. Students completed an extensive peer-evaluation battery at the end of the semester (Table 5.2). This battery captured their perceptions of conflict, satisfaction, and security with the team [196, 197, 195]. It can also assess behaviors like collaboration, contribution, and feedback [198]. Prior work shows that these instruments quantify aspects of social interactions that relate to performance [199, 200, 201, 202]. I used a participant's responses to these surveys to build a gold-standard baseline model to infer their final score.

The peer-evaluation contained the following validated instruments:

- *Team Conflict [196]* — Conflict represents the perception of incompatible goals or beliefs between individuals that cannot be trivially reconciled. This battery contains three scales, "task conflict", "process conflict", and "relationship conflict". When individuals perceive less conflict, it is associated with performance enhancement [199, 200]. This is likely because the positive outlook leads to better motivation [200] and satisfaction [201].

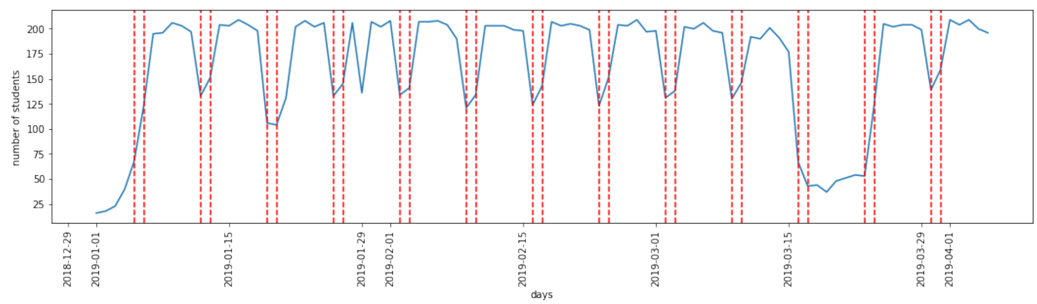


Figure 5.2: **Logs over time.** The number of connected students reduces during the spring break (week of 15th March) and with weekends (vertical red lines).

- *Team Satisfaction [197]* — Satisfaction reflects the contentment of an individual with their situation in terms of their expectations. Dissatisfaction with one's team can lead to lower levels of task performance [200, 201] and also moderate the effects of conflict on performance [202].

- *Psychological Safety [195]* — This construct captures a *"shared belief held by members of a team that the team is safe for interpersonal risk taking "* [195]. This is associated with individual learning progress as they are more amicable to experiments and feedback [195].

- *Team Member Effectiveness [198]* — This measure encompasses five dimensions [1]: (i) contributing to the project; (ii) interacting with collaborators; (iii) monitoring progress and providing feedback; (iv) expecting quality; and (v) relevant knowledge and skills. These characterize behaviors related to the individual-level construct, "team member effectiveness" [198].

Network Data

The dataset for this study is sampled from the large-scale dataset of WiFi AP logs described in section 3.2. The data spans a time frame of 95 days between January 1 2019 and April 5

[1]While the other scales were self-evaluations, this score is the average of how their peers evaluated a team member

2019 (Figure 5.2). On average, the time between the first log entry for any one of a participant's devices and the last was approximately 90 days. The logs indicated both the room and the building where the participant was connected to the WiFi. Along with coauthors, I manually categorized 204 buildings to best express the purpose of that space [108, 188] — for example, "academic", "dining", "green spaces", "recreation", and "residential". The raw logs were processed to obtain periods when students were dwelling and collocated (as described in subsection 3.2.2). Over the semester, the median collocation duration of a student with at least one other, was about 70hrs.

5.1.2 Feature Engineering

The low spatial resolution of the collocation makes it insufficient to assert from isolated instances if collocation of group members were connected to their performance. However, intuitively, observing multiple collocation events over the semester can approximate collocated interactions. For instance, members of the same group might collocate regularly at a specific type of building. Therefore, I engineered features that captured such patterns.

Feature Extraction

I extracted relevant information at a week-level based on various semantically labelled behaviors (Table 5.3). "Individual features" characterized behaviors which are not explicitly social, but could impact performance (e.g, attendance.). "Group features" captured the behaviors of individuals related to their group, such as time spent collocated with other group members. This dissociation of features helps provide discriminant validity and assert that coarse collocation-based features are not confounded by an individual's general behavior, such as the time spent in academic spaces.

I derived the individual features based on the lecture schedule and semantic labels for buildings. To craft the collocation features, I used the same information but compute them as both absolute duration and a relative percentage (of collocation time spent by all members of the group). The collocation features also incorporated the time of collocation events:

1. *Scheduled:* Groups reported their regular meetings in a free-form response field during enrollment (subsection 5.1.1). Responses typically indicated a primary building (e.g., learning commons) along with a potential backup (e.g., library). However, teams also expressed meetings could take place at undetermined locations on campus. Moreover, groups often provided multiple tentative meeting times and places for a week. To accommodate all possibilities, this feature described the collocations between group members that occurred during any of the reported periods.

2. *Class:* This described collocations with group members during class times. This is different from the attendance feature because it considered collocation outside the assigned lecture room. For instance, students in Part 2 were expected to meet during class time, and not necessarily in the scheduled room for the class. Based on student reports, Part 2 teams did not necessarily use all class times in a week for meetings.

3. *Other:* This is a catch-all bucket to capture all other ad-hoc collocations. Only 4 groups in our study reported interacting with group members for non-academic reasons (e.g., "lived together"). Since, improvement of social bonds is related to performance [44, 46], this category encompasses impromptu collocations, that could be motivated by course milestones but also represent other serendipitous situations.

Feature Processing

Raw week-level features were aggregated to derive features that describe collocation behavior. All the raw features I extracted (Table 5.3) from the data were computed at a week-level for 14 weeks—5×14 for *individual features* and $(9 \times 2) \times 14$ for *group features*. To reduce the feature space, I calculated summary features to describe the entire semester of the individual. Specifically, for each feature extracted at a week level, I computed the *median*, the *mean* and the *standard deviation* for the study period. In addition to these, I also computed

Table 5.3: Raw features derived from the collocation data at a weekly level

Type	Description	Spatial Variants			
		Any	Academic	Residential	Recreational
Individual Features					
Attendance	Present at lecture room during scheduled time	–	✓	–	–
Dwell	Time spent at a place while stationary	✓	✓	✓	✓
Collocation Features—*Measured as absolute duration and relative to the group*					
Scheduled	Time spent with group members during reported weekly meeting times	✓	✓	✓	✓
Class	Time spent with group members during class hours	–	✓	–	–
Other	Time spent with group members at other times	✓	✓	✓	✓

the *approximate entropy* of the feature per individual [203]. This aggregation reduced the overall feature count to 20 and 72 for individual and group features, respectively.

5.1.3 Training and Estimation

I built multiple regression models to investigate how the collocation-based features estimate final scores in comparison to survey-based peer evaluation scores.

Model Descriptions

M_{PE} denotes the model trained on peer-evaluation scores (subsubsection 5.1.1) based on the self-reported survey responses provided by the instructors. M_{iWF} refers to the model trained on individual features and M_{gWF} describes the model trained only features that represent collocation among group members — potentially describing collocated social interactions. I assessed the discriminant validity in predicting final course scores with each subset of features without confounding effects from other features. Furthermore, I developed combination models to comprehensively understand how a combination of automatically generated features estimate academic performance ($M_{iWF.gWF}$).

Estimators and Validation

I evaluated all models through a 5-fold cross-validation process ensuring that members of the same project group remain in the same fold folds. To estimate the target variable (the final score), for each model described, I trained a Linear Regressor [204] to represent linear relationships between features and a Decision Tree Regressor [205] for non-linear relationships. Additionally, we also train a Gradient Boost Regressor [206], i.e., an ensemble learner. To determine the relationship between model features and final scores, I measured the correlation between the predicted value and the actual values. For internal validation, I compared these models to a rudimentary baseline M_0, which always estimated the median of the target variable from the training set.

Feature Transformations and Selection

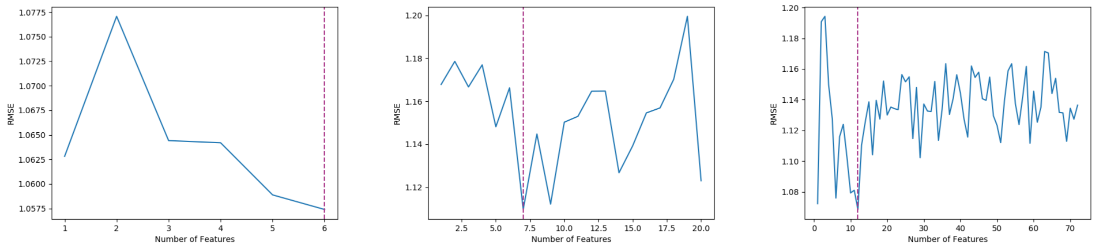
I performed the following transformations (fitted only on the training folds):

1. *Scaling Final Scores by Instructor* — Since the final score varies based on the instructor, I standardized the final scores based on the distribution of scores for each instructor in the training data.

2. *Impute Missing Data* — Some students had not completed all survey instruments. I imputed these missing values with the mean of the feature.

3. *Standardize the Features* — Converted to zero mean and unit variance [166].

4. *Mutual Information Regression* — We used the mutual information between the training features and the target variable for univariate feature selection [207]. The number of features selected were varied from 1 to k, where k was the total number of features in the model (Figure 5.5). We selected the k that minimized the RMSE (Root Mean Square Error) [208].



(a) M_{PE} (Linear Regression) (b) M_{iWF} (Gradient Boost) (c) M_{gWF} (Gradient Boosting)

Figure 5.3: **Mutual Information Feature Selection.** Number of features (X-axis) based on minimizing RMSE (Y-axis) with

Table 5.4: Model Performance. ('-':p<1, '.':p<0.1, '*':p<0.05, '**':p<0.01)

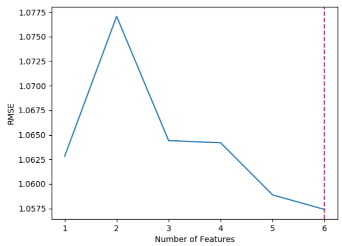
Model	Training Data	Estimator	*Pearson's R*	
M_{PE}	Peer Evaluation	LR	0.08	-
M_{iWF}	Individual Behavior	GB	0.14	.
M_{gWF}	Collocation Behavior	GB	0.24	**
$M_{iWF.gWF}$	Individual + Collocation	GB	0.25	**

5.1.4 Results

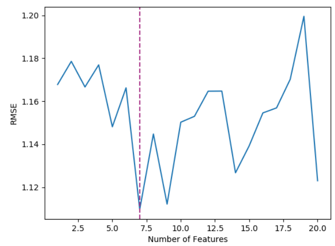
Model Comparison

Table 5.4 summarizes the results with the best estimator for each model. For any set of features, only the estimator that minimized the RMSE was considered for comparison between models.To compare models I used *Pearson's r* to describe the covariance of each model's estimate with the final scores of the students. This coefficient characterizes the complete association by considering all observations [209].

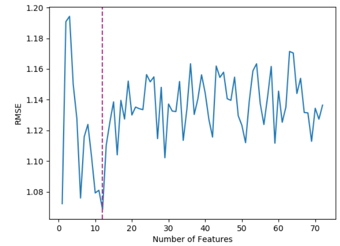
All models exhibited an improvement over M_0 — the rudimentary median estimator. None of the models based on peer evaluation features (M_{PE}) were found to be significant, but among them Linear Regression showed the most error reduction. For M_{iWF} the best estimator used Gradient Boost. Its estimates were more significant but with a weak correlation of 0.14. In comparison, for M_{gWF} the best estimator, which used Gradient Boost, exhibited a very significant correlation of 0.24. I also compared the dependent overlapping correlations [210] of M_{gWF} against M_{PE} and M_{iWF} (with a confidence-interval of 90%).



(a) M_{PE} (Linear Regression)

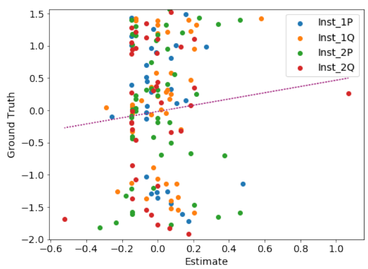


(b) M_{iWF} (Linear Regression)

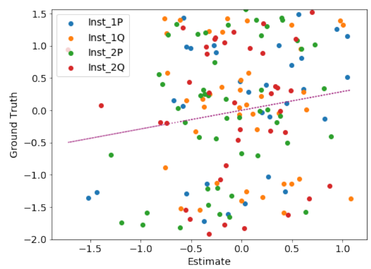


(c) M_{gWF} (Gradient Boost)

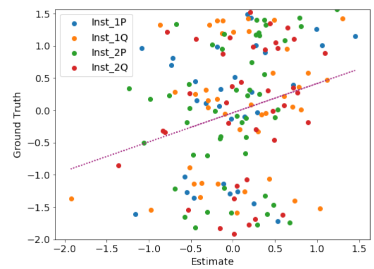
Figure 5.5: Best number of features based on mutual information regression



(a) M_{PE} (Linear Regression)



(b) M_{iWF} (Gradient Boost)



(c) M_{gWF} (Gradient Boost)

Figure 5.6: **Model Comparison.** Comparing the model estimates (X-axis) of an individual's final score (Y-axis); instructors are labeled by different colours

In both cases, the correlation of M_{gWF} with the final score was significantly different than that of M_{PE} ($p = 0.02$) and M_{iWF} ($p = 0.08$) (Figure 5.6). Additionally, incorporating both individual and within–group behaviors showed minor improvement. This improvement was not significant in comparison to M_{gWF}.

Interpretation

The results show that the model trained on students' collocation behaviors (M_{gWF}) outperformed the correlation of estimates obtained by modeling peer-evaluation and individual behaviors. While peer evaluation scores are expected to yield better correlations [200, 201, 45, 202], the social desirability bias in manually reporting team experiences can wash out the intricacies of actual team behavior [8, 9]. M_{iWF} was also found to be somewhat better than the peer-evaluation model. This already implies that dynamic offline behaviors have a significant relationship with academic performance. However, given the collaborative nature of the course in determining the final score of an individual, M_{iWF} falls short of M_{gWF}. These results indicate that even in collocation is important in projects that require agile coordination and collaborative work. This observation is in line with the concept of *spatiality*, where the presence of peers in the vicinity can affect individual performance even without direct communication [44].

The features in M_{gWF} aggregate collocation behaviors of students known to be socially connected over multiple weeks. The participants were expected to meet in person to work on their project towards their final score. In fact, a very small proportion of students reported collocation with team members for reasons unrelated to their project. Therefore, the fact that the collocation-based model (M_{gWF}) estimated the final score better than the dwelling-only model (M_{iWF}) provides evidence that inferring collocation of socially related individuals helps understand their performance. Moreover, $M_{iWF.gWF}$, which includes both group and individual behaviors, shows only a minor improvement over M_{gWF}. This further validates that it is indeed the collocation features that are predominantly explaining the individual performance in such group settings. Note, this passive inference does not explicitly discern what transpired during collocation incidents. However, it can be a complementary source of data that describe worker outcomes.

5.1.5 Summary

Collocation is known to be related to effectiveness of teams. Information workers tend to collocate, especially when working on projects. Therefore, one way to support their performance is to understand the extent to which collocation and performance are related. In this study, I demonstrated the feasibility of coarse collocation leveraged from WiFi network logs to investigate this phenomenon. I analyzed multiple project groups over a university semester to demonstrate how collocation behaviors of team members are related to individual performance.

5.2 A Study of Person–Organization Fit Through Latent Activity Routines

Individuals tend to thrive in organizations that share their values and beliefs [54, 50, 211]. Therefore. organizational studies have recognized this concept of person–organization congruence or "person–organization fit" as critically important [212, 213].

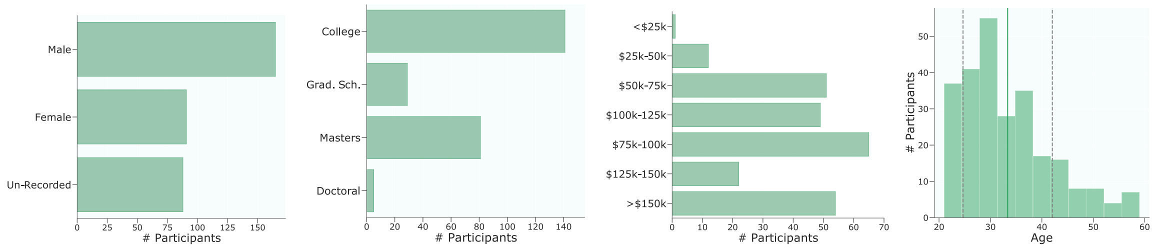
Past work on person–organization fit has often relied on static surveys, which are vulnerable to a variety of biases [66]. Often, these methods only represent an individual's perception of their values compared to an organization's values [56]. Therefore, the major limitation of such estimates of fit is their subjectivity [6]. In contrast, due to methodological constraints, objective measures of fit have only studied a single dimension of the employee (e.g., the level the organization values authority versus the level an individual values authority) [214]. These drawbacks prevent researchers from assessing more general ideas of fit, for which congruence is often a function of multiple dimensions.

On the other hand, sensors embedded in the environment can provide empirical estimates of workers within teams. I build on this idea by utilizing bluetooth beacons at the home and workspace to computationally infer individual routine patterns, and subsequently their similarity with the latent activity pattern of the organization [93]. I adopt this congruence of latent routines as a notion of P–O fit, or *routine fit*, and explore its relationship with measures of employee job performance and wellbeing. Through this study, I specifically address the following two research questions:

RQ1. What is the relationship between routine fit and different aspects of job performance?

RQ2. What is the relationship between routine fit and different aspects of wellbeing?

This study contributes to the literature in several key ways. First, I leverage passively sensed activity routines in cohorts to explore how they can be meaningfully used as person–organization variables and provide an objective measure of fit. Secondly, I go beyond



(a) Gender (b) Education (c) Income (d) Age

Figure 5.7: Summary of the participant demographic information.

single-occasion survey instruments and demonstrate the association of job outcomes with a data-driven representation of behaviors over time. Finally, these findings encourage future endeavors that incorporate the activities of cohorts to characterize their work experience.

5.2.1 Data

For this study I analyzed data acquired from the Tesserae Project section 3.1. I grouped individuals recruited from the same field site into "cohorts" (as they belong to the same organization or employer). Of the six cohorts, I selected four cohorts (each belonging to a different organization) of a sizeable sample to answer the research questions.

The models discussed in this study primarily involved data from the bluetooth beacons to build a model of routine fit. The beacons behave like access points that can be scanned by phone agent. Apart from this, data from the wearable is used to understand individual arousal levels – as a measure of wellbeing.

On entering the study, participants completed an initial battery to record details about demographics, job performance, personality and mental health traits via psychometrically validated survey instruments. To measure the daily fluctuations in these constructs, Ecological Momentary Assessments (EMAs) containing abbreviated versions of the initial ground truth instruments were disseminated periodically (Table 5.6).

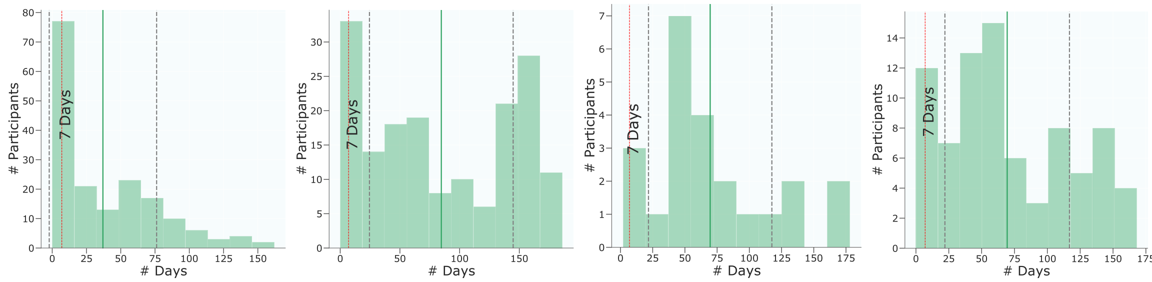
Table 5.5: 4 cohorts from the larger dataset were sampled for this paper. Each of these represents a unique field site tied to a specific organization.

Cohort	C_1	C_2	C_3	C_4
All Participants	294	177	26	89
Non-Colocated Beacons	176	168	23	81
After 7 Day Filter	113	139	20	71

Presence Sensing

To study the routine behaviors, I primarily processed the data from the Bluetooth beacons within an organization. Bluetooth beacon technology can approximate an individual's presence in its vicinity. Although it provides a coarse understanding of location it presents a tight accuracy radius, approximately 1-4 meters [215]. Unlike location sensing through mobile devices, which exposes an individual's every movement, presence sensing through beacons only relies on relative location, i.e., if they were near the beacon or not. Participants were asked to attach the static beacons to immobile objects at home and work. These objects essentially emit signals making them "observable" so that the participant's phone can discover them through periodic scans. It is not uncommon to employ bluetooth-like near field technologies to capture spatio-temporal data [216]. Additionally, bluetooth helps estimate indoor mobility and interactions [215, 217]. Dey *et al.* noted that individuals are at room level proximity to their phones (within 5-6 meters) for 90% of the day [217]. Thus, it is reasonable to consider the phone as a surrogate of the individual's presence.

Typically, the beacon designated for the home location was placed on the front-door and the one for work was situated on the individual's desk. An extended period of time away from either the home beacon or work beacon helps estimate when the individual left a particular place and entered another. Furthermore, the discontinuity in the presence of an individual near their desk indicates sessions in time that they are away from it. This could indicate casual breaks or scheduled meetings. An aggregation of these behaviors helps explain their routine [93].



(a) C_1 (b) C_2 (c) C_3 (d) C_4

Figure 5.8: The amount of data collected per participant varies based on when they enrolled as well as how compliant they were in terms of maintaining the beacons

Participants with home and work beacons located in the same place (based on beacon's GPS coordinates) were dropped from the study. These individuals were assumed to work from home, i.e, they do not find themselves colocated with their peers often. Other participants who were excluded had accidentally swapped their designated beacons. Individuals with less than 7 days of data were dropped as well. This decision helps maintain consistency with the self-report measures — some of which have a temporal resolution of a week. After filtering, the beacon data of 343 participants were analyzed (Table 5.5) to compute routine fit. Figure 5.7 summarizes the demographic information of the selected participants, and Figure 5.8 presents the amount of daily data provided by each cohort, with an overall average of 62.41 days of data for the selected participants. For each cohort, the majority of bluetooth data spans the months August, September, and October.

Note: In the dataset, 20% of the participants were "blinded" for external validation, i.e. their survey responses were obscured. Due to this, the job performance and wellbeing measures of only 249 (of the participants with adequate bluetooth) could be used for exploring specific relationships

Job Performance Measures

In this study, I evaluated three independent dimensions on which job performance can be described: task performance (IRB), organizational citizenship behaviors (OCB), and coun-

Table 5.6: The responses to the EMAs help determine the job performance and psychometric measures. *Garming Vivosmart 3* wearable supplied the arousal durations.

	Job Performance			Psychometric		Arousal Duration(s)	
	IRB	**OCB**	**CWB**	**Stress**	**Anxiety**	**Stressful**	**Restful**
Mean	42.81	6.85	1.14	1.97	1.69	20882	20275
Std	5.23	0.98	0.87	0.55	0.51	7694	6843
Max	49.00	8.00	5.99	3.37	3.36	41777	37480
Min	23.02	2.40	0.00	1.00	1.00	2775	1691
Scale	7-49	0-8	0-8	1-5	1-5	-	-

terproductive work behaviors (CWB) [32, 114]. Prior work has demonstrated personality congruence in an organization is related to work outcomes [218]. Hence, I used measures of personality based on the FFM as control variables.

Psychometric Characteristics

Beyond an employee's task accomplishment, measuring their general wellness is important to infer their success in an organization. P–O fit has looked at wellness from the perspective of satisfaction, commitment and propensity to leave, but there is little literature measuring mental health directly using the supplementary model of fit. Therefore, in this study I explore the relationship of P–O fit with self-reported anxiety, self-reported stress, and objectively measured high-arousal duration (via wrist wearable).

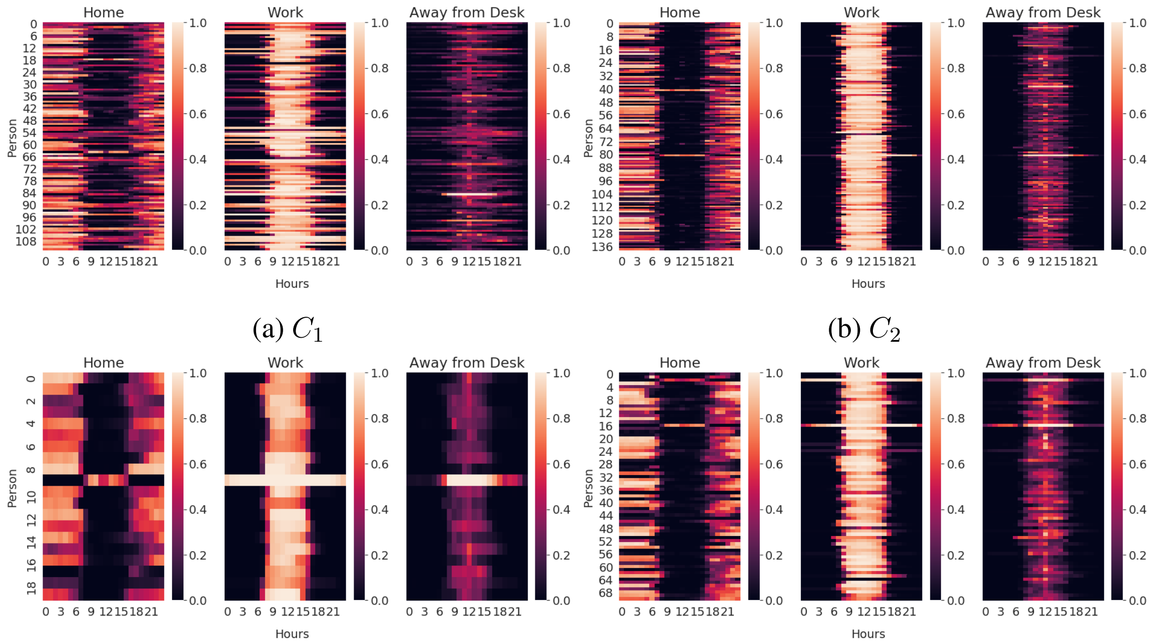
5.2.2 Methods

Aggregating Activities into Routines

Throughout the day an employee is engaged in a multitude of activities such as commuting, taking calls, creating slide decks and attending meetings. A routine is simply a sequence of such activities. I scope the concept of routine through an objective perspective of mobility. Using non-invasive bluetooth beacons I inferred the state of an individual – if they are at home, work or away from their desk (when at work). The temporal pattern of these states in a given day formed the routine for a day.

Operationalizing Individual Routines. The phone agent installed on participant smartphones periodically scanned the vicinity to locate other active bluetooth devices. Whenever a static beacon belonging to that individual was observed within a reasonable threshold of signal strength (-90 RSSI), the individual's presence at home or work could be determined. The instances at work were further deconstructed into sessions away from the desk – 5 contiguous minutes outside the range of the desk beacon is labeled as being "away". This data was chunked, or bucketed, in an hourly fashion to obtain the fraction of time at each hour an individual spends at *home*, at *work* and *away from desk* (when at work). The time at work represented a worker's habitual work hours, and the periods away from desk explained their internal schedules, such as meetings or breaks. The segments at home (and away from it) not only helps to infer commute times but also indicates spillover effects of work. Unlike the approach described in Eagle and Pentland which uses boolean representations, this method of fractional values per hour kept the information at a higher temporal resolution [93]. Using this method, each day was characterized by the 24–hour pattern of 3 different possible states the employee can be found in. This produced a 72 (= 24x3) dimensional vector that coarsely represents the routine for a given day. My study aims to characterize these routines in terms of individual mobility, or more accurately their presence near certain artifacts (front door/work desk), but this method can be extrapolated to any temporal activity.

Composing Organizational Routines. I constructed several routine vectors for each individual, proportional to each day they logged data in the study. The mean of these vectors represented the average routine of an individual. Practically, a third person observing the participant is most likely to see this behavior. A collection of individual routines belonging to the same organization would depict the "real" behavior or *observable cohort routine*. Figure 5.9 visualizes the aggregated behavior of each of the 4 cohorts. Even though, all the participants that were analyzed were primarily involved in information work, the organizations they work for are very different. C_1represents a large multinational company



(a) C_1 (b) C_2

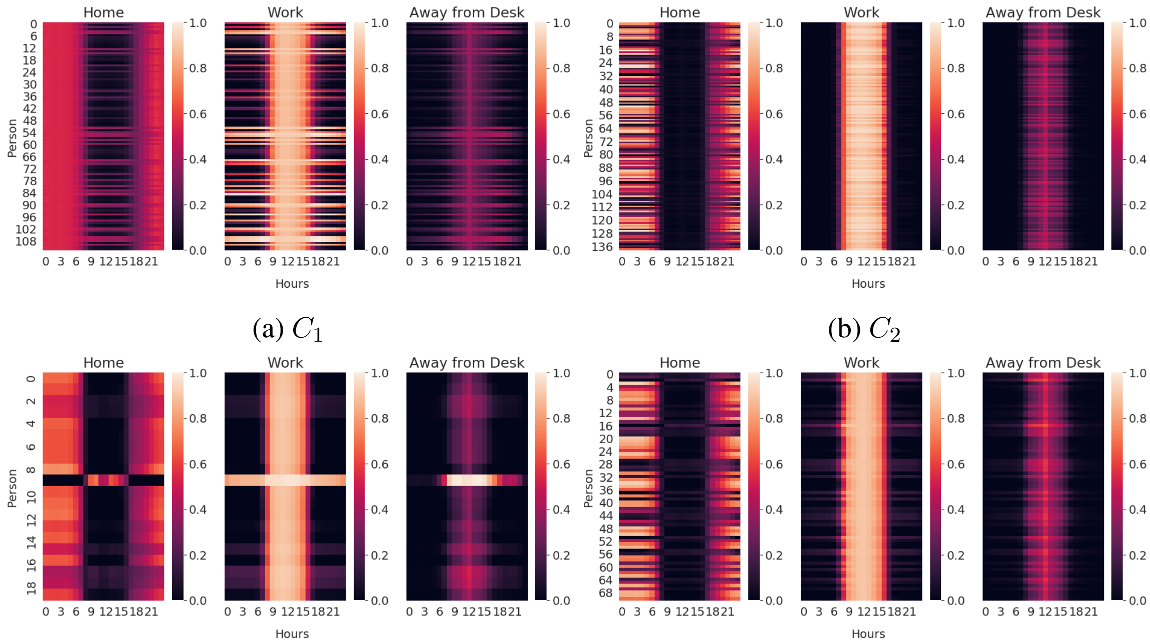
(c) C_3 (d) C_4

Figure 5.9: Every row of a heatmap illustrates the routine of an individual in an organization. Each column corresponds to an hour of the day. Brighter cells reflect individual presence – based on the beacon visibility.

primarily operating in the service sector. On the other hand, $C_2$is part of a manufacturing company that builds consumer products, C_3belongs to a small 50 people firm, and $C_4$is made of university staff. On eyeballing these it is quite evident that individuals in $C_2$and C_4demonstrated largely consistent routines. C_3showed regularity as well. Compared to these, members of C_1showed a lot more variation in the routines. This could be attributed to the fact that the $C_1$is a large consultancy where employees have diverse routines dictated by their specific client and project requirements. Hereon, whenever the section mentions an individual's routine, it is referring to the average of their daily routines.

Computing Person–Organization Routine Congruence

I adapted the eigen-decomposition method to aggregate behaviors of groups originally proposed by Eagle and Pentland [93]. Generally speaking, this technique identifies the primary patterns within data by assessing it in a latent space or "eigenspace" [219].

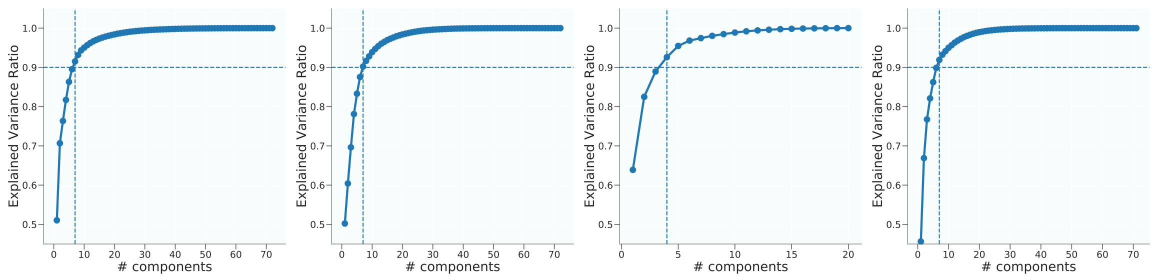


(a) C_1 (b) C_2

(c) C_3 (d) C_4

Figure 5.10: Latent organizational routines can be depicted using different sets of eigenvectors. Here the originally observed routines are projected on the most prominent eigenvector

Estimating Latent Routines. One can imagine taking a mean of all the average individual routines in an organization to quantify the routine of a cohort. However, that would only reflect how much time on average do employees spend at different places throughout the day, washing out the variance in the data and misrepresenting the behaviors of many employees. Thus arises a need to distinguish latent group behaviors that sufficiently represent the normative behavior for a given group. In order to do this, a *Principal Component Analysis* (PCA) needs to be performed. This identifies the *eigenvectors* or principal components of the observable cohort routine. These represent the most characteristic behavioral patterns shared by members of a cohort. However, these do not necessarily correspond to interpretable routines in themselves. Rather, these vectors reflect the underlying latent structure that empirically emerges from the patterns observed in the cohort. Any individual's routine can be practically expressed as a linear combination of these eigenvectors (Eq:Equation 5.1). Prior to this, the individual routines must first be mean-adjusted, i.e., an individual behavior should be contrasted from the mean activity of cohort. The mean

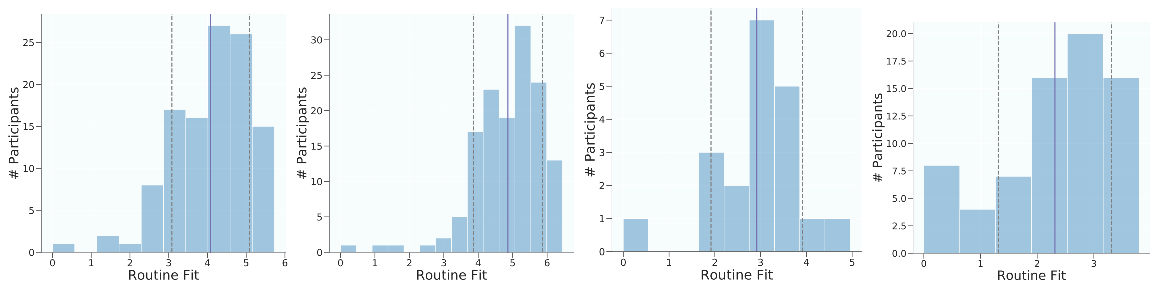


(a) C_1 (b) C_2 (c) C_3 (d) C_4

Figure 5.11: The cumulative variance explained by subsequent eigenvectors across different cohorts. We consider vectors that explain 90% variance to be reflective of normative behaviors

adjustment ensures that the primary principal component is independent to the mean of the data [220]. For any individual I, this adjusted routine will be referred to as $\Phi1(I)$. The PCA was applied on a collection of the mean-adjusted individual routines belonging to a cohort and eigenvectors are obtained. Figure 5.10 illustrates the cohort routines when they are expressed using only the most important eigenvector – the behavior that explains maximum variance in the cohort. Relative to the observed cohort routine (Figure 5.9), the projected routine (Figure 5.10) was able to highlight prominent behaviors. These represent the normative patterns within an organization. For e.g., every cohort shows a distinctly bright vertical column around 1200hrs in the “away from desk” block, indicating a commonly agreed upon lunchtime or regularly scheduled meeting that causes most employees to leave their desk.

Identifying Normative Routines and Explaining Behavioral Variance. Given that I described multiple employees in a high-dimensional space to explain routines, individuals could differ from each other in many different ways (across every hour on every feature). Furthermore, each individual can be compared to every other individual as well, implying a large set of comparisons. However, as already demonstrated earlier, a small set of latent patterns in behavior can explain a large part of the observed cohort routines. Figure 5.11 illustrates the cumulative variance explained by the eignvectors. Despite constructing the observed cohort routine with routines of multiple unique individuals, the PCA demonstrates



(a) C_1 (b) C_2 (c) C_3 (d) C_4

Figure 5.12: Distribution of ”Routine Fit” across different cohorts

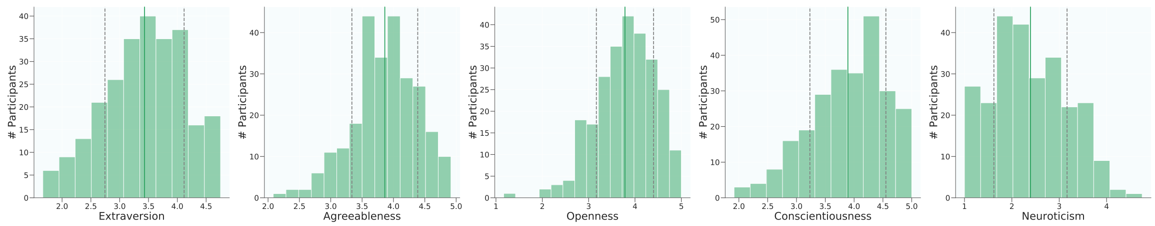
that the normative behaviors can be expressed with about 10 latent behaviors (explaining 90% of variance). Although challenging to interpret in their native form, each of these latent patterns is comparable to the actual routine such that they contain the same number of dimensions. The 10% of unexplained variance omitted from the normative patterns represents the irregular routine behaviors. In other words, these are the behaviors that are outside of the norm, accounting for individual differences within the population. For subsequent computations, I label the set of eigenvectors that explain these patterns as $U_1, U_2, ..., U_E$ (where E is the number of behaviors that explain 90% variance) – henceforth referred to as *latent cohort routines*.

Quantifying Routine Fit. The underlying latent cohort routines explain 90% of the observable cohort routine, but it is important to understand to what extent it can explain the routine of an individual within the organization. On projecting a single individual's routine onto the different latent cohort routines, I inferred the different weights corresponding to each behavior; referred to as $w_1, w_2, ..., w_E$. These weights denote the emphasis particular latent routines have in explaining the individual routine (mean-adjusted to $\Phi 1$). Having this information, for any given individual I, it is possible to reconstruct their activity routine based on the latent cohort routines as $\Phi 2(I)$. Essentially, Figure 5.10 illustrates $\Phi 2$ if it was computed with only the primary eigenvector. The subsequent analyses consider $\Phi 2$ computed with the first E vectors, which represents 90% of the observed routines. For individuals that behaved very similar to the norms of their organization their reconstructed routines and actual routines will be equivalent.

$$\Phi 2 = w_1 U_1 + w_2 U_2 + ... + w_E U_E \tag{5.1}$$

I conceive the *Routine Fit* $RF(I)$ of on individual (I) as the measure of similarity between the original activity routine of an individual, $\Phi 1(I)$, and the reconstructed activity routine $\Phi 2(I)$. For this, I first computed the Euclidean distance $D(I)$ between $\Phi 1(I)$ and $\Phi 2(I)$. At this step, it is important to note that routine fit only compares the congruence of routines within a cohort, but not across them because of how diverse they are (subsubsection 5.2.2). It is a relative measure because the absolute value of $D(I)$ is dependent on the normative routine of the cohort. As a result to control for inter-cohort differences in samples, these measures were standardized within the cohort as Z-Scores. Each of these standardized distances was subtracted from the cohort's maximum distance presenting a measure of similarity hereby called *routine fit*.

Figure 5.12 shows how the measures of routine fit vary across cohorts. Between $C_1$and C_2, the two cohorts of comparable sizes we observe that on average C_2has a higher routine fit than C_1. This insight supplements what we know from the observed cohort routine depicted in Figure 5.9. This distribution between the routine fit of $C_1$and $C_2$is also expected because these cohorts are part of organizations that operate differently. $C_1$is largely made up of consultants that work with individual clients and often have independent schedules. On the contrary, employees of $C_2$are engaged in research and development of consumer products and tend to rely on high internal collaboration. Therefore, the mean fit of $C_2$was bound to be higher than C_1because the employees of each cohort find themselves in different social contexts .



(a) Extraversion (b) Agreeableness (c) Openness (d) Conscientiousness (e) Neuroticism

Figure 5.13: Distribution of *Big Five* personality traits in the dataset

$$D(I) = \sqrt{\sum_{i}^{\mathcal{A}} (\Phi 1(I)_i - \Phi 2(I)_i)^2} \tag{5.2}$$

$$D(I) = ZScore(D(I)) \tag{5.3}$$

$$RF(I) = \max_{I \in C}(D(I)) - D(I) \tag{5.4}$$

Equation 5.4 represents my measure of routine fit and is analogous to the conceptualization of fit described by Edwards *et al.* in Equation 5.5 where the P is the person variable, E is the environment (or organization) variable and c is the theoretical maxima of fit within that organization [56].

$$F = c - |P - E| \tag{5.5}$$

In terms of my approach, P is equivalent to $\Phi 1$, reflecting the individual's routine as it is observed in the real world. E is equivalent to $\Phi 2$, describing the expected individual routine, given the normative patterns of their cohort.

Measuring Relationships with Routine Fit

After computing the routine fit of every individual, linear regression models were built to examine its monotonic relationships with each of the different outcome variables, Y [221]. These models included covariates for demographic information and intrinsic personality traits (Equation Equation 5.6). The different attributes of the FFM traits are correlated with different measures of job performance and mental health [34]. This data was collected during participant enrollment using the Big Five Inventory-2 (BFI-2) instrument (Figure 5.13) [222] – agreeableness (3.86 ± 0.53), conscientiousness (3.89 ± 0.66), extraversion (3.43 ± 0.69), neuroticism (2.39 ± 0.77), openness (3.78 ± 0.62). The demographics variables were chosen based on previous work [50, 212] – age (continuous), education level (ordinal), income (ordinal). None of the control variables were found to be significantly correlated with routine fit.

$$Y \sim age + education\_level + income + personality\_traits + routine\_fit \tag{5.6}$$

Additionally, I measured the *Variance Inflation Factor* (VIF) [161] for the covariates to check for multicollinearity among them. I performed this measurement iteratively for each covariate. At every successive step, the VIF of the covariates was found to be less than 1.4, which is far smaller than the conventional thresholds ($VIF = 5$ or 10) for excluding covariates. Therefore, the inflation of error caused by including these covariates in the model (Equation Equation 5.6) is inconsiderable.

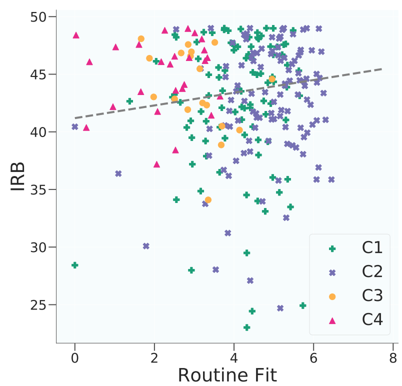
5.2.3 Results

RQ1: Routine Fit and Job Performance

The results of the linear regression show significant associations between routine fit and in-role behavior (IRB) as well as counter-productive work behaviors (CWB), as depicted in Table 5.7. There were no significant relationships found with the OCB test scores.

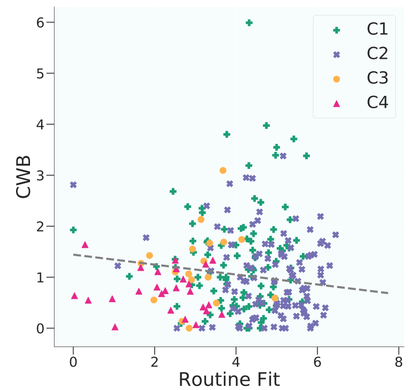
Positive correlation with In-Role Behavior. I found a positive correlation between an individual's routine fit and in-role behavior (Table 5.7a). Employees with home-work-desk patterns congruent to others in the organization tend to exhibit higher task performance. This is aligned with the Attraction-Selection-Atrrition theory that states workers who are congruent with their organizational patterns would be more likely to thrive as compared to

Table 5.7: Significant relationships between Routine Fit and job performance based on the linear model (Equation 5.6). Only significant covariates are reported. (. p¡0.1, * p¡0.05, ** p¡0.01, *** p¡0.001)



	Std. Coeff		
In-Role Behavior			
Openness	0.115	.	▲
Extraversion	0.119	.	▲
Agreeableness	0.204	**	▲
Conscientiousness	0.349	***	▲
Routine Fit	0.114	*	▲
$R^2 = 0.283$			

(a) IRB



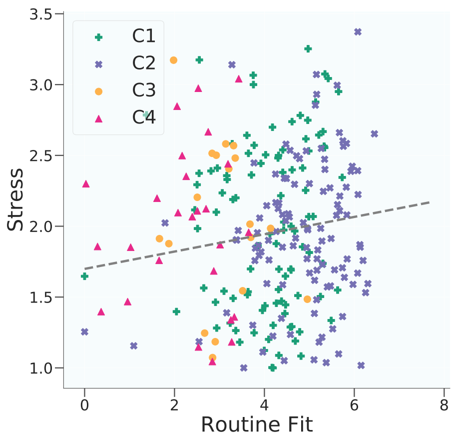
	Std. Coeff		
Counter-productive Work Behavior			
Age	-0.121	.	▼
Agreeableness	-0.138	**	▼
Conscientiousness	-0.249	***	▼
Routine Fit	-0.117	*	▼
$R^2 = 0.179$			

(b) CWB

those who are less congruent [223]. Moreover, this method shows a significant correlation with performance after controlling for the typical effects of personality [34]. One possible explanation for this lies in the conceptualization of organizational routine by Feldman and Pentland [224]. They describe routines as “sources of stability” that “encode organizational capabilities and knowledge”. Moreover, routine behavior of an employee is informed by organizational structure where macro-level changes only occur for the purposes of improved performance [224]. These effects are also grounded in the notion of entrainment — or the synchronization of routines — within the organization system. Syncing up with the task rhythm of a team is known to help increase coordination and task efficiency [225]. Individuals who are keyed into the dominant organizational tempos (i.e., have high routine fit) would exhibit greater task performance [226].

Negative correlation with Counter-productive Work Behavior. Next, I found a

Table 5.8: Significant relationships between Routine Fit and reported stress the linear model (Equation 5.6). Only significant covariates are reported
(. p¡0.1, * p¡0.05, ** p¡0.01, *** p¡0.001).



	Std. Coeff		
Reported Stress			
Neuroticism	0.351	***	▲
Routine Fit	0.113	.	▲
$R^2 = 0.186$			

negative correlation between an individual's routine fit and counterproductive work behavior (Table 5.7b). Congruence in home-work-desk patterns reflects a lower likelihood to be involved in deviant behaviors at the workplace. Non-conformity to the normative routines has been studied in other fields of psychology to understand its relationship with behavioral deviance [227]. From a social context, Bernburg and Thorlindsson studied the link between routines, differential social relationships and deviant behaviors [184]. In light of this, the negative relationship between routine fit and CWBs could indicate a lack of social connectedness between these workers.

RQ2: Routine Fit and Psychometric Characteristics

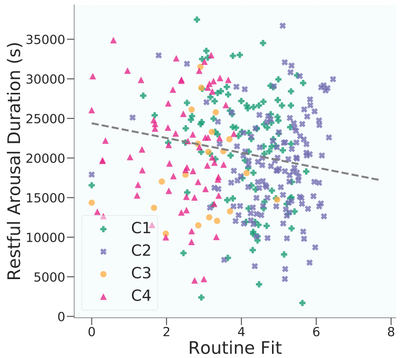
As per the results of the linear regression, used to model the relationship between routine fit and different wellbeing measures (Equation 5.6), I found that an individual's routine fit and reported stress were correlated. Additionally, I observed that routine fit is linked to their resting arousal and stressful arousal duration.

Positive Correlation with Reported Stress. Based on the results returned by the linear model, the routine fit showed a positive relationship with self-reported stress (Table 5.8). Before understanding this relationship, it is important to note the distribution of the self-report stress was skewed towards a lower score, reflective of low stress, with a mean of 1.97 on a scale of 1-5 with a max of 3.37 (Table 5.8). Most of the previous literature has only claimed relationships of fit and stress indirectly through other measures, such as strain and intent to leave. In fact, Arbour *et al.* found no significant association between stress and the congruence of behavioral norms in an organization[228]. Siegall and McDonald studied the relationship between value congruence and burnout, an extreme form of stress and found a strong negative correlation[229].

Given the ground truth data of our study does not capture the full range of the stress scale, there are little to none extremely high stress values reported. Another important detail to note is that the stress instrument used in this study does not delineate valence. With this in mind, it is possible that the individuals with higher stress reports were simply more involved in workplace activities [230]. In their study, Mark *et al.* state that, "People are happiest doing rote work and most stressed doing focused work"[26]. The interlinking between routine fit and stress could be indicative of high engagement work that is being performed by individuals following routines congruent to their peers. Moreover, the high routine fit could also represent a lack of autonomy in terms of work flexibility. When workers are not given sufficient agency to make decisions on task-related choices, including periods of work and schedule, they tend to be more stressed [231, 232]. In this regard a low routine fit could reflect resources being allocated to other aspects of life, such as one's social ties, subsequently reducing perceived stress [233, 234].

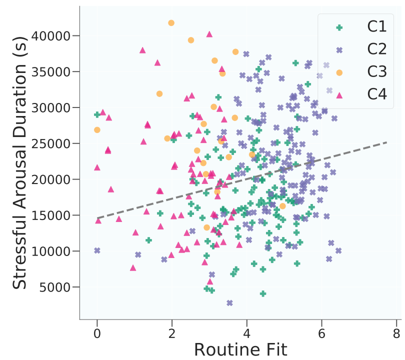
Positive Correlation with Stressful Arousal. On testing the relationship between routine fit and arousal duration, I observed a significant positive relationship with stressful arousal duration and a negative one with restful arousal duration (Table 5.9). Individuals with lower routine fit spend a longer amount of time in the restful state than those with a higher routine fit. On the other hand, the relationship of stressful arousal with routine fit indicates that high fit workers spend more time in higher arousal periods. This could either be indicative of physical activity or the response of an external stressors like a challenging task. Similar to the previous result, workers with high routine fit spending longer durations

Table 5.9: Significant relationships between Routine Fit and different arousal measures based on the linear model (Equation 5.6). Only significant covariates are reported (. p¡0.1, * p¡0.05, ** p¡0.01, *** p¡0.001).



	Std. Coeff		
Restful Arousal Duration			
Age	-0.186	**	▼
Agreeableness	-0.142	*	▼
Conscientiousness	0.258	***	▲
Routine Fit	-0.125	***	▼
$R^2 = 0.139$			

(a) Resting Arousal



	Std. Coeff		
Stressful Arousal Duration			
Conscientiousness	-0.206	**	▼
Routine Fit	0.151	*	▲
$R^2 = 0.107$			

(b) Stressful Arousal

in the stressful arousal state could be indicative of their involvement in engaging activities.

Post-Hoc Analyses

Controlling for Presence Duration

Recall that, the routine fit of an individual is computed based on the duration of their presence at different beacon locations. This method begs to question if the relationships between routine fit and the different dependent variables are simply the effect of an individual's time at home, work or desk. To untangle this, I tested the relationship of routine fit with our outcome variables by including these duration variables as covariates to the linear regression model:

$$Y \sim duration_{home} + duration_{work} + duration_{away\_from\_desk} + routine\_fit \tag{5.7}$$

Table 5.10: Significant relationships between Routine Fit and previously found significant relationships after controlling for durations (. p¡0.1, * p¡0.05, ** p¡0.01, *** p¡0.001).

	IRB	CWB	Reported Stress	Resting Arousal	Stressful Arousal
Home	***	**	.		***
Work	.		*		
Away from Desk	.	*			*
Routine Fit	***	***	***	**	***
R^2	0.95	0.90	0.90	0.85	0.81

The results of the regression depicted in Table 5.10 showed that the relationship between the calculated measure of routine fit with different job performance variables and psychometric constructs, still hold in models controlling for the duration variables. This is because routine fit models similarity in patterns across 72 different features in a high dimensional space, capturing information that cannot be acquired by matching 3 dimensions of duration [93]. This finding provides further evidence that routine fit expands upon simplistic measurements of duration to explain the outcome variables.

Generalizability of Routine Fit across Cohorts. As described in subsection 5.2.1, all of the cohorts are similar in that they represent employees involved in information work. Having said that, these cohorts vary in many different aspects such as organization size, workforce diversity, and company culture. Therefore, it is reasonable to argue that routine fit has a different relationship to performance and wellbeing outcomes for different cohorts. Even though subsubsection 5.2.2 describes the use of Z-scores to standardize routine fit, I only adjusted for observable differences and not the subjective differences between cohorts. This motivates me to empirically validate the effect of different cohorts by testing a random-effects model:

$$Y \sim age+education\_level+income+personality\_traits+routine\_fit+(1+routine\_fit|cohort) \tag{5.8}$$

Equation Equation 5.8 extends on the linear model used in our analysis (equation Equation 5.6 by incorporating a random term, $(1+routine\_fit|cohort)$. This random term tests if the slope of routine fit and the outcome variable (Y) varies across cohorts. I applied this model to find that routine fit does not vary significantly across cohorts for most of the outcome variables except for stressful arousal, for which the variation is only weakly significant ($p = 0.084$). This indicates that, even though the cohorts vary in nature, the findings presented in subsubsection 5.2.3 and subsubsection 5.2.3 are consistent across them.

5.2.4 Summary

Unlike previous constructs of person–organization fit, I used a bluetooth based passive sensing framework to conceive a new construct of P–O Fit, known as *routine fit*. Routine fit was measured by characterizing a worker's home–work routines on the basis of multiple dimensions. My findings resembled those from traditional methods of fit that also rely on value-congruence between workers and teams. I found routine fit was associated with high task performance and low workplace deviance. However, I also found that routine fit exhibited a positive correlation with perceived stress and arousal, implying high engagement work or possibly loss in autonomy when workers are highly coordinated. These findings support the use of passive sensing for teams as a means to disentangle new notions of worker socialization.

CHAPTER 6
PASSIVE SENSING FRAMEWORKS TO INFORM ORGANIZATIONAL DECISIONS

An organization is a collective of workers in a system that is held together by a common purpose. In my research, any reference to an organization is synonymous to companies or employers. While information workers can be free-lancers and self-employed, my focus is on information workers that find themselves employed by an organization or function within the purview of one. Organizations can be of varying sizes where workers are interconnected through multiple hops in social ties and stacks of hierarchy. In such a collective, individual behaviors are in a constant osmosis with others [12, 186]. Therefore, information workers need to adjust to these social norms to sustain a healthy work experience. As discussed in subsection 2.2.3, these norms can emerge as an aggregation of perceptions, but they can also be the result of purposeful organization wide decisions.

Simply by the nature of its complexity, organizational behavior is nebulous to understand. It is challenging to capture by surveying small samples and even more difficult to assess regularly. In turn, making organization wide changes are hard leading to stagnant cultures and inability to cope with unforeseen events such as a pandemic. Through the previous chapters, I have already demonstrated that passive sensing frameworks can indeed describe individuals and teams by overcoming limitations of traditional assessments. To holistically support an information worker, I extend a new research question that considers the role of larger organizational entities on a worker's experiences.

RQ III: *How can passive sensing frameworks inform organizational change?*

In this chapter, I discuss two investigations to answer this research question. First, I modeled perceptions of large organizations by analyzing self-initiated accounts to describe normative practices within a community that explain individual worker performance. Second, I modeled behaviors of large communities over long periods of time to provide a flexible behavior-based toolkit to inform proactive and practical community interventions for resumption of work during COVID-19.

Together, these studies demonstrate opportunities for passive sensing frameworks to: (i) dynamically characterize organizational culture by reflecting the views of a large community of workers, (ii) flexibly complement existing organizational decision–making processes for crisis response by modeling naturalistic behaviors.

6.1 Characterizing Organizational Culture with Passively Collected Accounts of Workplace Experiences

Organizational culture (OC) refers to certain norms and principles that are believed to optimize the workforce [235, 50]. It embodies a core value system which affects the development and execution of new ideas, and the management of unexpected events like crises [236, 237]. While metrics such as revenue and profit are standard methods to gauge the effectiveness of an organization, the culture of an organization is a key indicator [60]. From the employee's perspective, OC is related to their loyalty and commitment [237].

Organizational studies have employed a variety of survey instruments to quantify OC [59, 238, 239, 64, 240], which are limited in scalability and temporal granularity. Besides, the data in these studies lacks reliability because information workers are often concerned about the confidentiality of their opinions [10, 11]. Therefore, the workplace context can invite multiple biases, such as response (or non-response) bias, study demand characteristics, and social desirability bias [66].

In contrast, workplace review platforms contain self-initiated and anonymous reports [241] that stand to mitigate many of the biases introduced by survey studies [242]. Glassdoor is one such platform with publicly posted reviews of workplace experiences. Not only do these reviews contain objective information like pay, hours and benefits but also the free-form text that encapsulates various nuances of OC [235, 243]. The language in this shared experience reflects an organizational culture where recognition is not prioritized but concern for others and cooperation is upheld. In fact, through the affordance of descriptive text, platforms like Glassdoor provide an accessible, scalable and flexible medium to express cultural and ecological differences [244].

My work leverages the language used in publicly visible employee reviews to computationally model OC and augment our understanding of it. Specifically, this study has the following research aims:

Aim 1. *To operationalize OC as a multi-dimensional construct and validate it with language on Glassdoor.*

Aim 2. *To computationally model OC of an organizational sector, and evaluate if it explains employee job performance.*

My first research aim strives to build a usable construct of OC, based on Glassdoor data, that captures various aspects like interpersonal relationships, work values, and structural job characteristics. Towards this, I used established frameworks from the domain of organizational psychology [59, 238, 239, 64, 240] to identify job descriptors related to OC and represented them in the lexico-semantic space of word embeddings [245]. This produced a codebook of lexical phrases that closely align with different dimensions of OC.

Next, given a reliable representation of OC I examined if it explained information worker performance [50, 237]. I applied the computational OC construct on Glassdoor reviews and quantified the OC of various companies by sector (e.g., management, production, or computer). On a ground truth dataset from the Tesserae project (section 3.1) I incorporated OC to find that it improved on intrinsic traits (such as demographics and personality) to explain an employee's task performance and citizenship behavior. Therefore, I present empirical evidence that OC can explain human functioning.

Table 6.1: Example paraphrased excerpts in *Pros* and *Cons*.

Pros	**Cons**
1) Great teams 2) Talented co-workers 3) Not stressful 4) Good work-life balance	Most departments offer no flexibility in work schedule. My manager doesn't allow me breaks for doctor appointments, child's school activities
Good work environment, nice people. Lots of fun working on cool technology. Location is also superb.	No communication from upper management, Pay is not nearly as competitive as market salaries.
Friendly, outgoing coworkers. Very health-conscious environment. Activities are encouraged and supported.	Little recognition for overtime hours, no WFH alternatives even with bad weather, poor work-life balance

6.1.1 Using Glassdoor as a Passive Sensing Framework for Employee Experience

In this study, I leveraged crowd-contributed workplace experiences from Glassdoor to validate a computationally operationalized framework of OC (Aim 1), and to quantify the OC in an organization of information workers (Aim 2).

Glassdoor is an online platform (launched in 2008), for current and former employees to write reviews about their workplace experience. As of 2018, it hosted 57M individual accounts and 35M reviews posted for 770K companies [246].

Glassdoor reviews require ratings and free-form text. Employees can rate their overall experience on a scale of 1 to 5, and optionally add ratings for fields like career opportunities, compensation, and senior management. The free-form text field requires employees to submit descriptons of their workplace experience, in separate sections for *Pros* and *Cons*. This text describes many salient workplace themes, such as work-life balance, management, pay, benefits, growth opportunities, facilities, and interpersonal relationships. Table 6.1 shows three example excerpts in *Pros* and *Cons* components.

Glassdoor strives to be a trusted and transparent platform for job searching [242]. Both contributing content and consuming content necessitates an individual login. It only allows individual accounts with *permanent, active email address, or a valid social networking account* to submit content, with a maximum allowance of *one review, per employee, per year, per review type* [247]. Glassdoor moderation involves proprietary content-analysis

Category	Organizational Culture Dimensions
Interests	Conventional, Enterprising, Social
Work Values	Relationships, Support, Achievement, Independence, Recognition, Working Conditions
Wk. Activities	Assisting & Caring for Others, Establishing & Maintaining Relationships, Guiding & Motivating Subordinates, Monitoring & Controlling Resources, Training & Teaching Others, Coaching & Developing Others, Developing & Building Teams, Resolving Conflicts & Negotiating
Social Skills	Instructing, Service Orientation
Struct. Job Characteristics*	Consequence of Error, Importance of Being Exact, Level of Competition, Work Schedules, Frequency of Decision Making, Freedom to Make Decisions, Structured versus Unstructured Work
Work Styles	Concern for Others, Leadership, Social Orientation, Independence, Integrity, Stress Tolerance, Self Control, Adaptability, Cooperation, Initiative, Achievement
Interpersonal Relationships*	Frequency of Conflict Situations, Face-to-Face Discussions, Responsibility for Outcomes & Results, Work w. Group or Team

Table 6.2: 41 Org. descriptors from O*Net to represent the dimensions of OC. The category column indicates the O*Net category of the descriptors. Categories with '*' are subcategories within the "Work Context" cateogry.

technology as well as human moderators. Any reviews deemed to be incentivized or coerced, are either not allowed or removed from the platform. To ensure a non-polarized distribution of reviews, Glassdoor implements a key incentive policy known as, "give to get" [241]. In this model to get full access to all reviews, viewers must contribute their own review. This paradigm diminishes the effect of reactionary reviews by self-selected users by encouraging more neutral opinions by all types of users [248]. The content posted on Glassdoor remains anonymous, and the moderation strategies ensure that no sort of individual-identifiable detail is disclosed in the content.

6.1.2 Aim 1: Operationalizing Organizational Culture

To measure OC through Glassdoor reviews, I first operationalized it with language over a three-step approach: 1) Identify descriptions of multiple dimensions of OC. 2) Transform the descriptions into word-vectors to represent OC within a linguistic and semantic context, and 3) Compare the word-vector based OC construct to filter Glassdoor posts related to OC and qualitatively investigate the posts' keywords to establish face-validity.

Identifying Descriptors of Organizational Culture

Language used by a community (or organization) provides a unique lens to interpret its culture [235, 244]. To understand the extent to which a text expresses OC, I first established an ontology of job aspects that are indicative of different OC dimensions. For this, I obtained job aspect descriptors from the Occupational Information Network (O*Net). O*Net (*onetonline.org*) is an online database of occupational information developed under the sponsorship of the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA). These descriptors were motivated by organizational research literature [249, 250, 251]. O*Net describes 189 different job descriptors, categorized in 17 sub-categories, which are further grouped into 8 primary categories. Each of the 189 job descriptors, like *Stress Tolerance*, *Level of Competition* and *Independence*, is accompanied by a description.

To identify descriptors that align with known dimensions of OC, I along with a coauthor independently inspected each of the 189 descriptors in O*Net on the basis of four OC instruments, *Organization Cultural Inventory* [252]), *Organization Culture Profile* [50]), *Hofstede's Organization Culture Questionnaire* [64], and *Organization Culture Survey* [239]). Any discrepancies ($n = 23$) was mutually resolved by both authors on agreeable themes and concepts. Overall this procedure had a Cohen's κ (inter-rater reliability) score of 0.89 This process retained 41 descriptors, each of which describes an aspect of OC (see Table 6.2). Also note that these dimensions are not necessarily mutually exclusive or disjoint [50, 253], leading to a significant overlap in ensuing analysis.

Transforming Descriptors into an OC Construct

Simply tokenizing the keywords in the 41 descriptors of OC would not adequately capture the concept of OC. Therefore, I encapsulate the linguistic and semantic context of these descriptions by using word embeddings [254, 255]. This approach represents words as a vector in a higher dimensional space, where contextually similar words tend to have vectors that are closer.

Measure	Total	Mean	Stdev.
Reviews	616,605	6,702	8312
Pros Sntncs.	1,386,787	15,073.77	18,408.64
Pros Words	10,747,265	17.42	20.91
Cons Sntncs.	1,715,875	18,650.82	22,786.10
Cons Words	17,150,342	27.81	47.24

Table 6.3: Descriptive stats. of Glassdoor dataset of 92 companies (sourced from top 100 of Fortune 500). Aggregated values are per company.

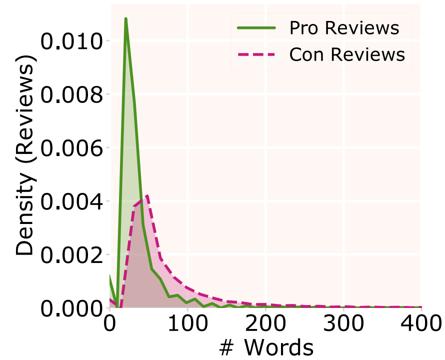


Figure 6.1: Dist. of no. of words per review in the Glassdoor dataset of Fortune 100 companies.

I used pre-trained word embeddings in 50-dimensions (GloVe: trained on word–word co-occurrences in a Wikipedia corpus of 6B tokens [245]). Building on prior work of representing job aspects in lexico-semantic dimensions [256], I transformed the explanations for each of the 41 descriptors (Table 6.2) into a 50-dimensional word-embedding vector.

These 41 word-embedding vectors essentially characterize multiple dimensions of OC in a latent semantic space. Collectively, they constituted my operationalized construct of OC in the form of a 41-D word vector.

Validating our Operationalization of OC

To establish the validity of operationalized OC for practical use. I qualitatively inspected the top keywords from our Glassdoor dataset, which were relevant to OC.

Compiling the Glassdoor Dataset. To obtain a diverse but voluminous dataset on Glassdoor, I consulted the *Fortune 500* list (ranked by revenue) [257] and obtained the top 100 ranked companies. Only 8 of these companies appear in the list of *Fortune 100 Best Companies to Work For* [258]. Therefore, it is reasonable to assume that my sample was not dominated by companies with positively-skewed employee experiences. I obtained the public reviews of these organizations by web scraping Glassdoor. For each review, I collected the textual components (segregated into *Pros* and *Cons*) and the reviewer's employment information — role and location (Table 6.3 and Table 6.1). Note that the content

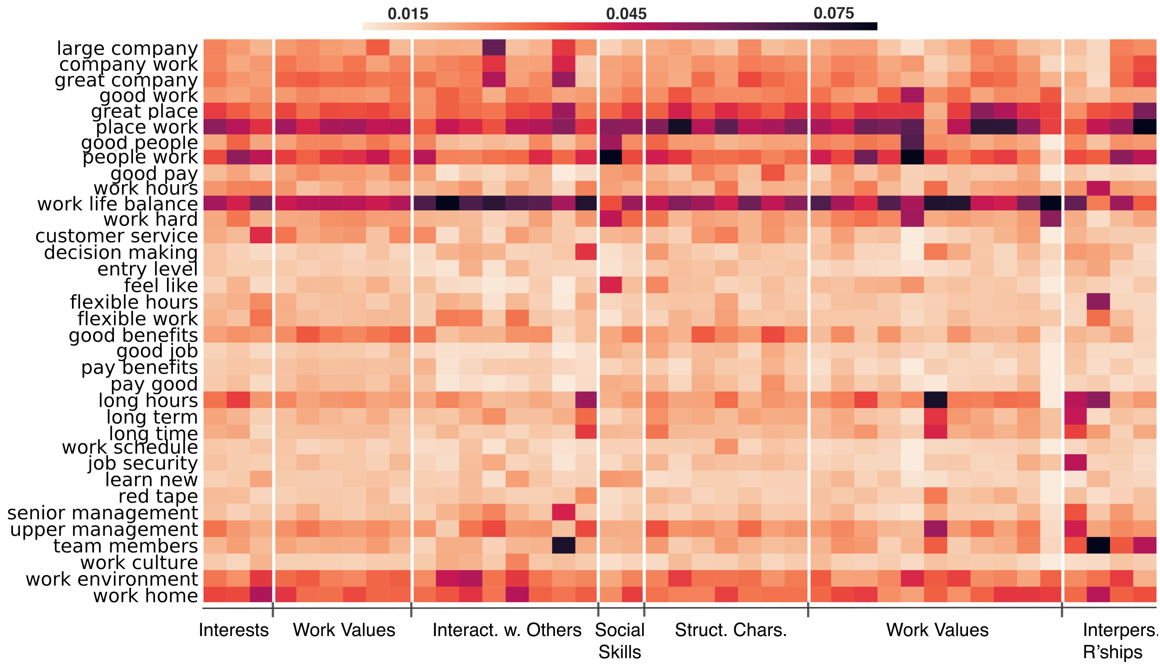


Figure 6.2: Top n-grams in sentences about OC (excluding lexical variants of keywords). Darker colors (higher TF-IDF score) indicate greater relative importance within a particular dimension. Dimensions have been categorized corresponding to the scheme in Table 6.2

distribution is skewed towards the *Cons*, but this observation aligns with activity on other review platforms [259].

Filtering Posts about Organizational Culture First, Id derived a word-vector representation of every sentence in the 616,605 posts (~3M) from the Glassdoor dataset. Next, I used cosine similarity to measure the similarity between each sentence's word-embedding representation and each of the 41 dimensions of OC [260, 24]. Higher cosine similarity indicates that the sentence is semantically similar, or "talks about" that particular dimension of OC. I retained any sentence that exhibited a similarity of more than 0.90 with any of the OC dimensions [79]. Table 6.4 enlists a few paraphrased examples.

Establishing Face and Construct Validity

I obtained the top 100 keywords (n-grams, n=2,3,4) in all the retained sentences and computed the TF-IDF score for these keywords across each of the 41 OC dimensions (similar to [255]). This score describes the relative importance of each keyword in the sentences (Figure 6.2). I drew on the validity theory [261], to establish face and construct validity

Example Text	OC Dimension
Great training, really genuine and supportive colleagues, great ways to get involved with interest groups— Proposal writing, research for new industry areas, volunteer activities	Social
In many instances rank was invoked just to prove a point, rather that using data for the same.	Importance of Being Exact
The drive to succeed is key, however, it's not a cut throat competition - people are humble and people at all levels are interested and willing to develop those at the lower career levels.	Level of Competition
If you have a goal and willing to work on it, senior management will have a genuine interest in helping you succeed.	Coaching and Developing Others
A lot of emphasis is on firm activities making it difficult to build relationships as you can only meet coworkers on Fridays, if they do come.	Establishing and Maintaining Interpersonal Relationships
New recruits are immediately given responsibility, and can take complete charge of their career development.	Initiative
Lot of group work makes the work easier and more fun.	Independence

Table 6.4: The word-vector representation of these sentences that show a cosine similarity of 0.90 or greater for the corresponding OC dimension. Note that the same sentence can reflect multiple dimensions, but I only list one for brevity.

of contextualizing OC in Glassdoor data by qualitatively examining the importance of the keywords in the OC dimensions. The most dominant keyword across several dimensions was **work life balance**, and its lexical variants like "life balance", "work life". This recurrence could be because notions of work–life balance has many facets (beyond work-family conflict) such as personal needs, social needs and team work [262]. Other keywords were more discriminatory between OC dimensions. For instance, the keyword **"good benefits"** was most important in reviews about dimensions like *Support* and *Recognition*. For employees, reward systems garner reciprocal loyalty and increase the perceived organizational support [263]. Another such keyword is **"job security"**, which is most relevant to experiences that refer to the *Frequency of Conflict Situations* dimension. This draws from the fact that employees in workplaces that have high disagreements require more security and stability of employment [64]. Other examples of identifiable n-grams are **"flexible hours"** or **"flexible work"**. These keywords gain maximum importance in text associated with the *Face-to-Face Discussions* dimension. Prior research found that teams with fluid hours accommodate more interactions [264]. Similarly, the terms **“long hours”** and **“long time”** are important in texts related to the *Stress Tolerance*. Longer working hours not only causes fatigue but also increases an employee’s exposure to work-related stressors [265, 266, 267].

Together, this evidence indicated that the OC construct built from curated O*Net job aspect descriptors was able to capture the OC-related language used in Glassdoor reviews.

6.1.3 Aim 2: Modeling OC and examining its Relationship with Job Performance

Prior work in the domain states that organizational culture (OC) influences individual performance in the workplace [63, 62]. Therefore, to evaluate if my operationalization of OC is meaningful, I modeled it with the performance of information workers across different occupational sectors.

Compiling the Groundtuth Dataset

I obtained groundtruth dataset of individual job performance from three companies C_1, C_2, and C_3 — from the larger Tesserae Dataset (section 3.1), and the Glassdoor reviews of these three companies. Together, it included individual difference attributes and job performance of 341 information workers across 18 unique sectors in three companies C_1, C_2, and C_3 in the U.S (Table 6.5). The individual attributes included demographic details such as age, gender, education, supervisory role (supervisor / non-supervisor), income, and their role in the organization. Additionally, it also included the measurement of personality traits (FFM) and scores from a Shipley scale [268] that measured the executive function in terms of fluid and crystallized intelligence.

In terms of job performance, the dataset provides two measures— 1) *the IRB scale* [117] measures their In-Role Behavior that characterizes the proficiency at performing appointed activities and tasks, and 2) *the OCB scale* [124] measures Organizational Citizenship Behavior which characterizes participation in extra-role activities that are not typically rewarded by the management [118, 119, 120, 121]) .

Table 6.5: Summary of individual attributes for Aim 2.

Measure	Scale	Range	Mean	Stdev.	Distribution
		Independent Variables			
Demographics					
Age		21-64	34.15	9.01	
Gender	Categorical: Male — Female				
Job characteristics					
Tenure	Ordinal: 10 values [¡1Y,1Y,..¿8Y]				
Supervisory Role	Categorical - IT — Non IT				
Personality Traits (BFI-2)					
Extraversion	1-5	1.67-4.91	3.42	0.68	
Agreeableness	1-5	2.08-4.91	3.85	0.54	
Conscientiousness	1-5	1.92-5.00	3.90	0.65	
Neuroticism	1-5	1.00-4.67	2.44	0.75	
Openness	1-5	1.17-4.91	3.79	0.60	
Executive Function (Shipley)					
Crystallized: Abs.	0-25	0-23	17.11	2.97	
Fluid: Voc.	0-40	0-40	33.06	3.93	
		Dependent Variables			
Job Performance					
IRB	7-49	20-49	44.48	4.57	
OCB	20-100	32-100	56.20	10.28	

The dataset categorized participants into 18 unique sectors based on role such as, "business and financial operations" (115), "computer and mathematical" (105) and "management" (50). Other participants were in sectors such as "office and administrative support" and "healthcare practitioner". Put together, I studied 25 combinations of company and sector (eg. $\{C_1$, Computer and Mathematical$\}$, $\{C_2$, Management and Consultancy$\}$, etc.).

Accordingly, I obtained $23,791$ reviews on Glassdoor (22,794 for C_1, 574 for C_2, and 423 for C_3). These reviews contained 1,654, 134, and 108 unique roles respectively. I mapped these roles to 18 sectors using their semantic similarity (using pre-trained word vectors trained on 6B tokens on the entire Wikipedia corpus) [245], and next, two researchers manually verified the mapping, and edited the sector label wherever necessary.

Modeling and Quantifying OC by Org. Sector

I first collated all the reviews posted in different company sectors. Then, using word-embedding based cosine similarity, I obtained the similarities of every review sentences

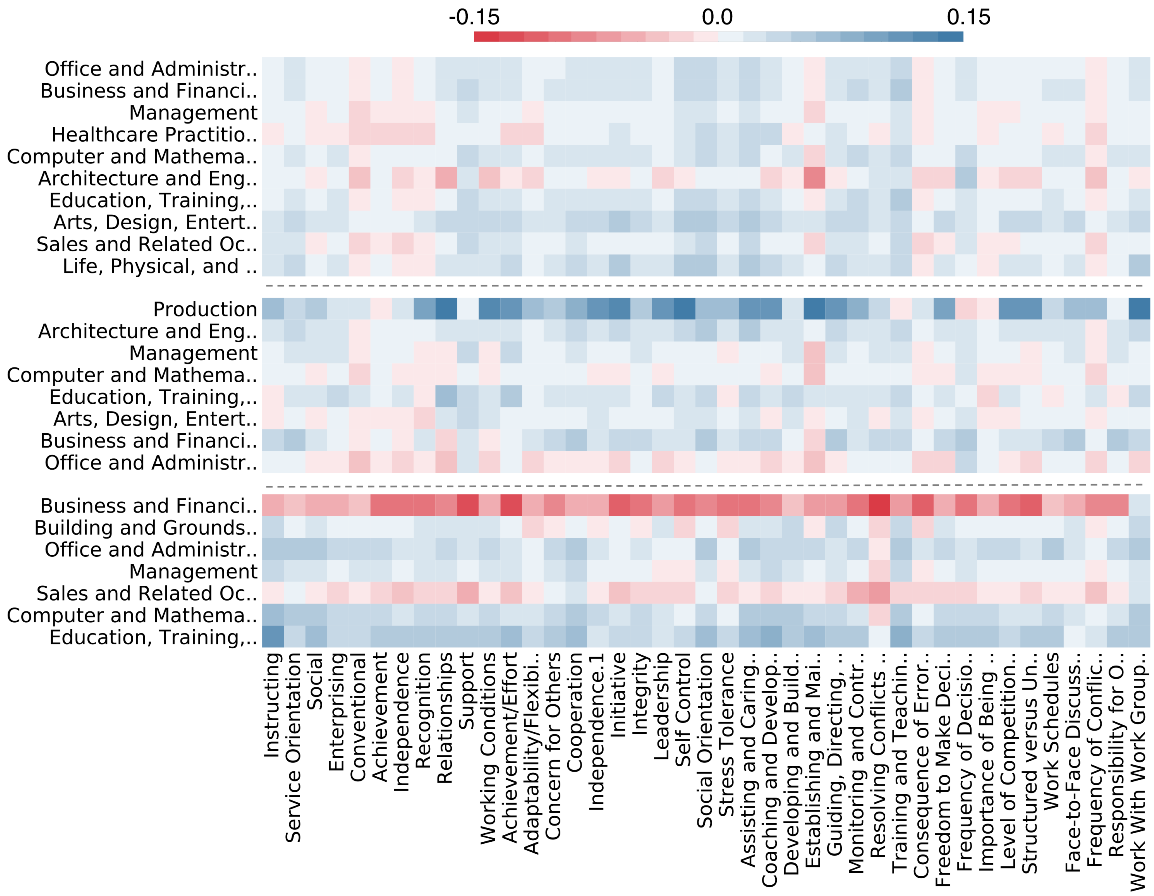


Figure 6.3: Organizational culture as quantified via Glassdoor data per organizational sector in three companies C_1 (top), C_2 (middle), and C_3 (bottom). The color and intensity of the cells represent the positivity or negativity in that dimension of organizational culture.

with each of the 41 OC dimensions. Since every dimension of OC can be valued differently by each employee, I qualified the raw similarity score with the help of Glassdoor's *Pros* and *Cons* structure. I assigned a weight of $+1$ to those sentences labeled as *Pros* and -1 to those sentences labeled as *Cons*. I obtain the weighted average of cosine similarities for each dimension. Together, this reflected a 41-dimensional vector of OC. The value per dimension is equivalent to how positive or negative that dimension was lexico-semantically referred to in an organization's Glassdoor reviews.

Figure 6.3 shows the distribution organizational culture in 41 OC dimensions in the groundtruth dataset. OC was observed to vary across sectors both within and between companies. For example, the reviews from employees in the sector "business and financial operations" shows contrasting trends — while the reviews in C_1 and C_2 talk about OC in a similar way, the reviews of C_3 typically discuss dimensions of OC in *Cons*.

Relationship between OC and Job Performance

I hypothesized that our approach of operationalizing OC can explain an individual's job performance which I obtain at the beginning of this section [237, 50, 62].

> **Hypothesis.** Organizational culture provides significant explanatory power towards one's job performance.

To test this hypothesis, I first built a baseline model (*Model 1*), with individual attributes, to predict job performance (Equation 6.1). This was motivated by prior work that has established that individual difference attributes (such as personality and executive function) are strong indicators of job performance [34, 269, 270, 271, 35, 272]. Next, I built an experimental model (*Model 2*), where I incorporated OC alongside the individual difference variables, and predicted the same job performance measures again (Equation 6.2).

I used three types of linear regression models with regularization, Lasso (L1 regularization), Ridge (L2 regularization), and Elastic Net (both L1 and L2 regularization), and two non-linear regression models, Support Vector Machines (SVM) regressor and Random Forest regressor. To tune the parameters of the models, I used a grid search [273]. For validation, I used a leave-one-out (loo) methodology to train and predict over the dataset. Finally, I collated all the predicted data points, and obtain the pooled model performance measures — these include Pearson's correlation and Symmetric Mean Absolute Percentage Error (SMAPE) to evaluate the predictive accuracy of our models, and R^2 to evaluate the model fit (here JP is job performance).

$$JP \sim gender + age + income + supervisory\_role + tenure + exec.\_function + personality \tag{6.1}$$

Table 6.6: Summary statistics of the "best" regression models in *Model 1* and *Model 2*, where *Model 2* includes organizational culture, whereas *Model 1* does not. ***: p <0.0001

Measure	**IRB**		**OCB**	
	Model 1	*Model 2*	*Model 1*	*Model 2*
Algorithm	Lasso	Ridge	Ridge	Ridge
R^2	0.23***	0.28***	0.15***	0.24***
Pearson's r	0.43***	0.45***	0.32***	0.41***
SMAPE	3.67	3.65	6.96	6.71

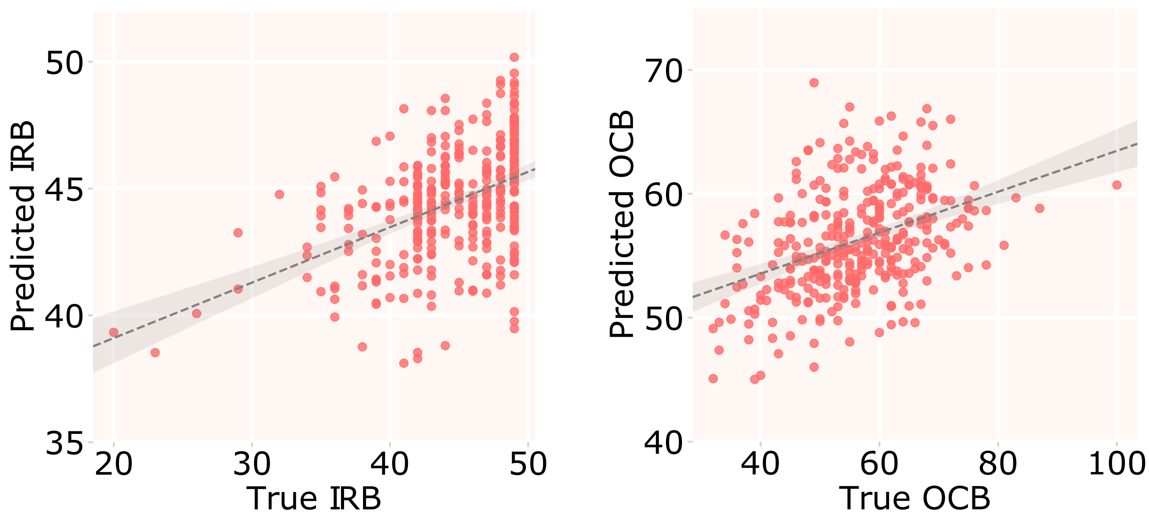


Figure 6.4: Scatter plots showing true and predicted values per *Model 2* of IRB (left) and OCB (right).

$$JP \sim gender+age+income+supervisory\_role+tenure+exec.\_function+personality+OC \tag{6.2}$$

Does OC Explain Job Performance?

Model Performance. Table 6.6 summarizes the fit and accuracy metrics of *Model 1* and *Model 2* for predicting job performance measures (IRB and OCB) (see Figure 6.4 for scatter plots). I found that *Model 2*, which included the OC as a feature, performed better than the *Model 1*. In the case of IRB, *Model 2* fit 22% better, and *Model 2* predicted 5% better on the pooled correlation, with 0.6% lower SMAPE. In the case of OCB, the model showed 60% better fit, 28% better prediction correlation, and 4% lower predicted error. All these models fit and predict with statistical significance ($p < 0.01$).

Table 6.7: Summary of regression coefficients in predicting job performance by *Model 2*. This reports top 10 coefficients ranked on variable importance [276].

IRB		OCB	
Variable	**Coeff.**	**Variable**	**Coeff.**
Freq. of Conflict Situations	0.59	Adaptability/Flexibility	-49.92
Service Orientation	6.31	Work Schedules	1.45
Recognition	24.10	Face to Face Discussions	0.36
Independence	-9.93	Importance of Being Exact	-0.46
Responsibility for outcomes	0.89	Coaching Others	-37.43
Working Conditions	-8.58	Instructing	-36.56
Freq. Decision Making	-10.80	Wk. w/ Work Group	-0.003
Enterprising	0.96	Conventional	-167.92
Monitoring Resources	0.80	Support	-72.41
Initiative	-9.20	Maintain Relationships	75.35

Model Validity. Furthermore, I verified if this improvement was due to random noise created by additional features in *Model 2*. Therefore, to reject the null hypothesis that a randomly generated 41-D vector will perform better than *Model 1*, I drew on permutation test approaches [274, 275]. I ran 10,000 permutations of randomly generated OC vectors — completely unrelated to actual review data or any other real data. I found that the probability (p-value) of improvement by a randomly generated feature set was 0.002 for IRB and 0.0001 for OCB. Thus, I rejected the null hypothesis and established statistical significance in the observed improvement with addition of OC based on a sector's Glassdoor posts. Table 6.7 reports coefficients of the top 10 OC dimensions (as ranked using variable importance [276]) in *Model 2*.

Therefore, supporting our hypothesis in the previous subsection, I find that OC as computationally modeled using Glassdoor reviews per organizational culture explains job performance of individuals in those organizational sectors.

Post-Hoc: Does Language tell us more than Ratings?

Finally, after establishing that quantifying OC with Glassdoor posts of a sector *does* significantly explain individual performance at workplace, I revisitd the question, "is quantifying via language actually effective?" As Glassdoor is a platform that allows employees to provide ratings, it was important to examine if features based on the language offer anything

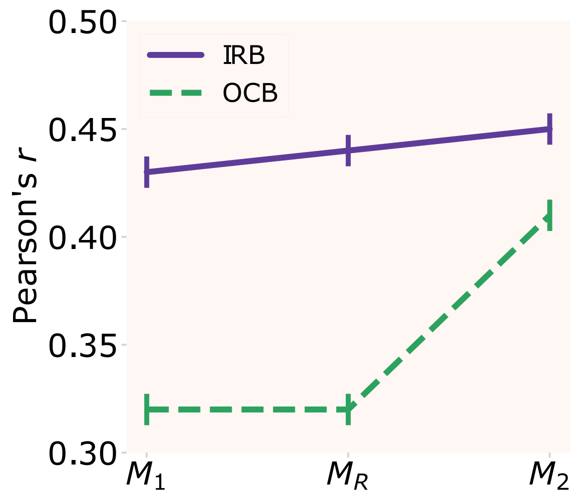


Figure 6.5: Pearson's r of Models predicting individual job performance (M_1: Model w/o OC, M_R: Model w/ Org. Sector wise Rating, M_2: Model w/ OC via Language)

more than raw scores. For this, I built a third model where I only replace OC in Equation 6.2 with mean aggregated rating per sector. The Ridge model performed the best in both the job performance measure predictions; in case of IRB, this model shows an adjusted R^2=0.24, Pearson's r=0.43, and SMAPE=3.65. In case of OCB, this model shows Adj. R^2=0.14, Pearson's r=0.32, and SMAPE=6.95 – this model performs only as good as *Model 1* (Table 6.5). So, Glassdoor content when quantified in the lexico-semantic space bears greater explanatory power in comparison to a single numeric rating. These results further validated my approach of operationalizing OC as a multi-dimensional construct [50] instead of relying on a single value.

6.1.4 Summary

This study empirically studied OC with a passive sensing framework that leveraged large-scale employee-contributed workplace experiences posted on Glassdoor. I examined the linguistic dynamics in public-facing anonymized company reviews to describe culture and developed a theoretically-grounded rendition of it in the form of a codebook. Subsequently, I developed a lexicon to encapsulate culture in terms of 41 dimensions. I illustrated a natural-language methodology to model culture for organizational facets (such as sector) and tested its explanatory power in predicting employee performance.

6.2 Modeling Organizational Networks to Aid Infectious Disease Crisis Response

In the wake of the Coronavirus Disease (COVID-19) [67], the U.S. witnessed more than half a million cases at universities [277]. Colleges were one of the many organizations that had to decide how to resume operations while also controlling spread of an infectious disease [278, 279]. To reduce on-campus infections and the likelihood of superspreading events, a recommended form of non-pharmaceutical intervention (NPI) is partial closure of the campus [68]. The advancement in teleconferencing technology equips organizations to continue operations by adopting a form of campus closure — in universities this is through remote instruction (RI) [280]. In higher education, this population of students, staff, and faculty is functionally similar to information worker. However, when an organization keeps large portions of its population away due to closure, it can lead to broad, negative, and indiscriminate impact on the community. For individuals, this lead loss in learning outcomes [71, 72] and decreased mental wellbeing [73, 74]. For the larger organization, it can lead to losses in auxiliary revenue (e.g., boarding, parking, dining, etc.) [69, 70]. Therefore, by simply relying on RI, organizational campuses struggle to balance community health with the demands of learning, economy, and broad wellbeing [281]. Instead, there is a need for a more versatile approach to design closure policies that empowers policymakers to accurately assess impact of closure interventions and model more data-driven targeted intervention strategies.

This study showcases a new approach that organizations can take to design closure policies by leveraging data from their existing WiFi infrastructure. My methodology, *WiFi mobility models* (WIMOB), involves constructing anonymized mobility networks of campus (Figure 6.6a), which helps determine extended periods of collocation — or "proximate contact" [282]— between individuals to describe contact networks on campus. Mobility has always been used to dynamically model disease spread of influenza [283], rubella [284]

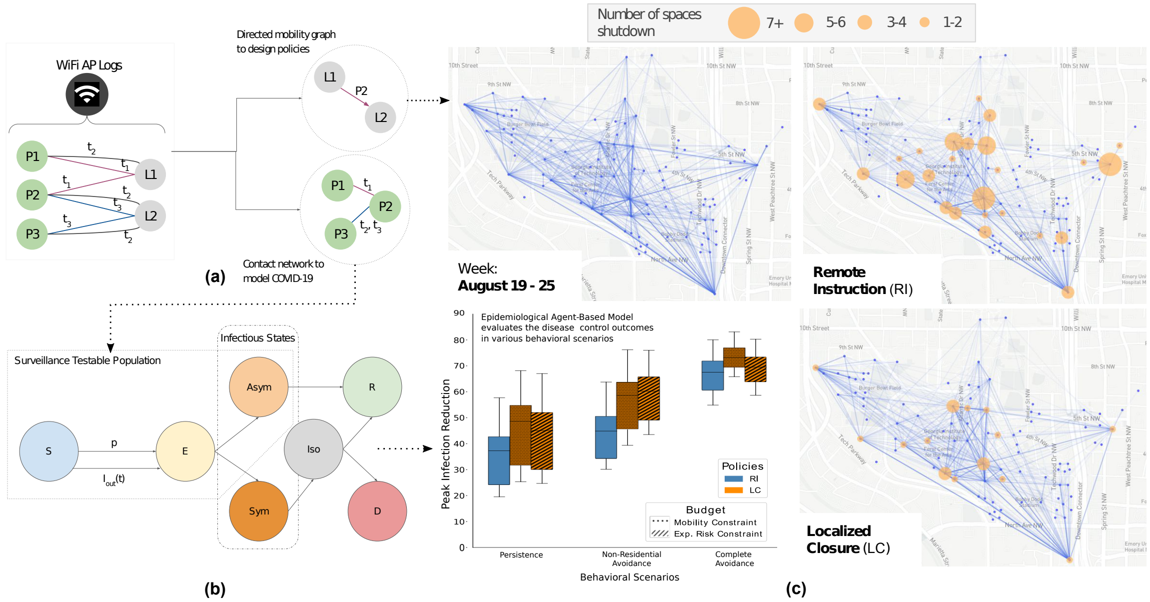


Figure 6.6: The WiFi mobility models (WIMOB) methodology uses anonymized network logs to model campus mobility and target spaces for localized closures (LC) (a) WiFi network logs reflect timestamps when people's devices associate with access points (APs) on campus. WIMOB mines these logs to characterize mobility as a bipartite graph that describes people (e.g., $P1$, $P2$) visiting campus locations (e.g., $L1$, $L2$) during different times (e.g., t_1, t_2). Since people's devices can proxy their presence, I estimate collocation (e.g., $P1$ and $P2$ were collocated at $L1$ at t_1), and movement ($P2$ dwelled at $L1$ and then at $L2$). (b) I use the collocation network construct a SEIR–based epidemiological ABM, calibrated to Fall 2020 incidence of COVID-19 (c) WIMOB highlights mobility behavior to evaluate and inform policy. (c)*–top-left:* Mobility on campus between the top 100 most frequented locations on the GT campus in the Fall semester of 2019. Edges only connect points of significant dwelling and thus do not represent pedestrian routes. (c)*–top-right:* RI is a form of broad closure which affects a large number of students and locations.(c)*–bottom-right:* By contrast, I propose to use WIMOB to parsimoniously identify a small set of spreader locations within buildings and design LC policies. (c)*– bottom-left:* I use our epidemiological ABM to evaluate these policies under different budgetary constraints and various behavioral scenarios (Persistence, Non-Residential Avoidance, Complete Avoidance). Our study shows that LC policies provide equal or better control on the disease spread, and yet minimize the burden on campus compared to RI.

and COVID-19 [285, 286] showing the effectiveness of mobility restrictions at a regional–, or city–level [287, 288, 289, 290, 291]. These studies typically rely on cell tower localization or aggregating GPS information from mobile phones [292]. Neither of these data sources are easy to access for organizational campuses. Alternatively, other studies have relied on specialized user-facing data logging applications (e.g., contact tracing mobile apps) [105, 106, 107, 293], but these require mass adoption and maintenance to represent the entire community longitudinally. In contrast, my work repurposes already existing managed WiFi networks to model mobility, which provides room level granularity for mobility [108, 109, 15, 110] and consequently indicates proximate contact [282]. This passive sensing framework does not need any additional surveillance infrastructure. With the appropriate privacy considerations, an organizational campus can obtain such data at a low cost, continuously and unobtrusively.

Particularly, WiMob enables a more expressive toolkit for university policymakers that represents contact longitudinally and allows them to assess closure at the granularity of a room, suite, or hall. Thus, it is a passive sensing framework that lends itself to the design of targeted interventions for localized closures within an organization (LC). I demonstrate the utility of WiMob with data collected over two years, of approximately $40,000$ anonymous occupants and visitors of GT, a large urban campus in the U.S. — including about $16,000$ undergraduate students, $9,000$ graduate students, and $7,600$ staff members. First, my results found WiMob to reveal novel structural characteristics of community network. Second, by constructing and simulating an agent-based epidemiological model (ABM) over the people–people contact networks (Figure 6.6b) derived from the collocation identified with WiMob (Figure 6.6a), I illustrate the design of better LC policies which control disease spread while minimizing disruption to community.

6.2.1 Leveraging Managed WiFi Networks

I present two sets of analyses in our work. The first set contrasts structural characteristics of contact networks described by WiMob with current practices that use enrollment data (En). In the next set, I used WiMob to build an epidemiological model (an agent-based model over the contact networks, referred to as ABM) and analyze the remote instruction (RI) and localized closure (LC) interventions in terms of their differences in dynamic disease-control outcomes and burdens to campus.

WiFi Mobility

To characterize mobility (WiMob), I utilized the dataset of WiFi logs as described in section 3.2. I had access to logs from Fall 2019 through Fall 2020. The logs indicated if any of $40,000$ unique visitors were connected to one of $6,959$ the WiFi access point (AP) across all of campus. This was limited to indoor spaces where APs are located and the scope of this localization is at the granularity of a room or suite [108, 149]).

Studies on RI policies tend to assume that contact in universities is largely informed by EN — transcripts showing student course registration [294]. Therefore, to compare the efficacy of WIMOB, I refer to aggregated insights of enrollment networks (based on course registration transcripts for GT). Note, I did not cross-identify any students. Additionally, I used publicly accessible course schedules to approximate schedules of de-identified individuals and inferred if they were students or staff, and non-residential or residential.

Like most universities, GT's managed WiFi network is not equipped with any Real-Time Location System (RTLS) [150, 151]. RTLS systems use Received Signal Strength Indicator (RSSI) values from multiple neighboring APs to provide high precise localization of individuals in terms of time and space. However, deploying such systems requires surveying the entire network. Additionally, precision localization raises more privacy concerns. These factors together make it challenging for universities to justify the deployment of RTLS, unlike small retail settings that can monetize RTLS insights directly (e.g., insights on footfall can be tied to improving revenue).

Contact and Movement Networks. WIMOB leverages the logs to create bipartite graphs K_t, for each day t, which connect P users to L access point locations (Figure 6.6a). Any edge, $\{p, l\}_i$ indicates the i^{th} instance when a p was dwelling at l. These edges describe the time period of dwelling. Subsequently, by comparing all edges in K_t we can infer if different individuals are collocated near an AP to create a contact network, G_t, for each day t — between any collocated $p_i, p_j \in P$. These networks define the contact structure for an epidemiological agent-based model at every time-step. Similarly, by inspecting the sequence of dwelling locations for any p in graph K, I computed a mobility network, H_t — between locations $l \in L$. My approach considered collocation as a form of *proximate contact* — people in the same room — and therefore established collocation only when this occurred for "an extended period" [282] of approximately 40 minutes.

Modeling Policies. In this study, I compared the disease outcomes and burdens of 2 policies, Remote Instruction (RI) and Localized Closure (LC), both of which were modeled with WIMOB. For RI I inferred enrollment size of each course in Fall 2019 by determining the number of unique individuals that visited lecture locations during scheduled times. After the first week, I applied the RI by removing all visiting edges in K_t for any $l_c \in L_{\mathsf{RI}}$ if visits were during lecture times of course c with an enrollment ≥ 30. This helped create counterfactual contact networks G'_t. The removal of edges from K described the mobility budget of RI and the structure of G'_t indicated the risk of exposure budget. I designed LC with these budgets by identifying locations for closure (L_{LC}) with different algorithms, such as *PageRank [295], Eigenvector Centrality [296], Load Centrality [297], and Betweenness Centrality [298].* When a location was closed, I removed all edges in K_t connected to any $l_x \in L_{\mathsf{LC}}$. I aggregated the movement graph H_t over a week and applied the algorithms to identify locations. Subsequently, I identified the number of top-ranked locations to remove such that the resultant counterfactual contact network G''_t has is within 1% of the budget. The budgets varied for different behavioral scenarios and I only compared policies within the same scenario.

Disease Simulation

Here I summarize our epidemiological model and calibration process.

Agent-Based Model. I constructed an agent-based model (ABM) that captures the spread of COVID-19 between individuals active on campus. This ABM leveraged the contact networks produced by WIMOB. The simulation iterated a time-step each day with the underlying contact networks i.e., G_t for no interventions, G'_t for RI, and G''_t for LC. Each agent in the ABM followed a modified version of *susceptible–exposed–infectious–removed* (SEIR) template that disambiguates the *infectious* compartment into *asymptomatic* and *symptomatic*. New infections were introduced to the model either externally or internally. External transmission arose because individuals could contract the virus outside campus and bring the infection back for local spread [299, 280]. I adopted data of positive cases

Table 6.8: Model Parameters of the ABM

Parameter	Definition	Value	Std	Source
p	Transmission probability: For any edge between a *susceptible* and *infectious* individual in the contact network, p is the probability that the *susceptible* person will enter into the *exposed* state. This only dictates internal transmission	0.034	0.007	Calibration
α	Scaling factor of the normalized confirmed cases in the surrounding county. This is the parameter for us to generate $I_{out}(t)$	0.032	0.0032	Calibration
I_0	Proportion of population that is *asymptomatic* at day 0	0.012	0.0009	Calibration
p_S	Probability of *exposed* persons becoming symptomatic	0.66	-	[302]
Δ_S	Incubation period (days) since the first day of exposure	5	-	[302]
$\Delta_{\text{Asym}\to R}$	Asymptomatic duration (days); it is the time taken for an *asymptomatic* person to recover since the first day of exposure	7	-	[302]
Δ_I, σ_I	Time of a *symptomatic* entering *isolated* since the first day of exposure of a *symptomatic* person	8	2	[303]
Δ_R, σ_R	Time for recovery for a *symptomatic*, since the first day of exposure	12	2	[304]
p_D	Death rate under isolation	0.0006	-	[304]

The variables p, α, and I_0 are estimated by calibrating the simulation model on the first 5 weeks of positivity rates provided by GT surveillance for Fall 2020, while incorporating external cases from Fulton County. These parameters were found by validating the ABM on the remaining weeks of Fall 2020.

from Fulton county [300] with a scaling factor α to estimate the probability that a *susceptible* individual, who is active on campus, was infected from interactions that took place outside campus. Internal transmissions were determined by p, as the probability of *susceptible* individuals in contact with an *infectious* one. I calibrated the parameters related to disease transmission by training and validating our models on the positivity rate reported by GT surveillance testing [301].

Calibration. For this study, I estimated three key parameters: (i) infectious individuals at day 0, (ii) transmission probability between infectious and susceptible individuals, and (iii) the probability of infection transmission from contacts outside the network. I estimated the range of optimal parameters for disease transmission by minimizing the root means square error (r.m.s.e) between the Georgia Tech surveillance testing positive rates [305, 301] and the observed positivity rate of the model every week— percentage of new *asymptomatic* cases out of the total testable population. The surveillance testing conducted by GT was designed for detecting individuals who contracted COVID-19 without showing Flu-like symptoms within the community [305]. I calibrated the model on the positivity rates on the first 5 weeks of Fall 2020. To attain a point estimation of the optimal parameters, I fitted the model to predict trends in the remaining weeks by running a numerical optimization algorithm, Nelder-Mead [306]. To account for quantitative uncertainty, I estimated a range of parameters, within 40% of optimum r.m.s.e [288]. For other model parameters, I adopted values proposed by previous studies on similar populations [302, 303, 304]. Table 6.8 shows a full list of our parameters.

Variation of Parameters. Note that the calibration characterized latent factors associated with pandemic-related cautious behaviors, including the relationship with external transmission. And these factors could be related to "county characteristics, partisanship, media consumption, and racial and ethnic composition" [307]. To account for the effect of these varying latent factors on disease outcomes, I performed additional calibrations for hypothetical variations in disease spread. For these analyses I kept the GT mobility behavior constant while calibrating the model on different time periods of surveillance testing and on positivity rates of different U.S. universities — University of Illinois at Urbana-Champaign [308] and University of California [**–ucbcv19testing˝**], Berkeley. I evaluated RI and LC on these variations and describe the design of these complementary experiments for additional robustness. in SI Sensitivity Analyses. Results of all calibration parameters are described in Table 6.8.

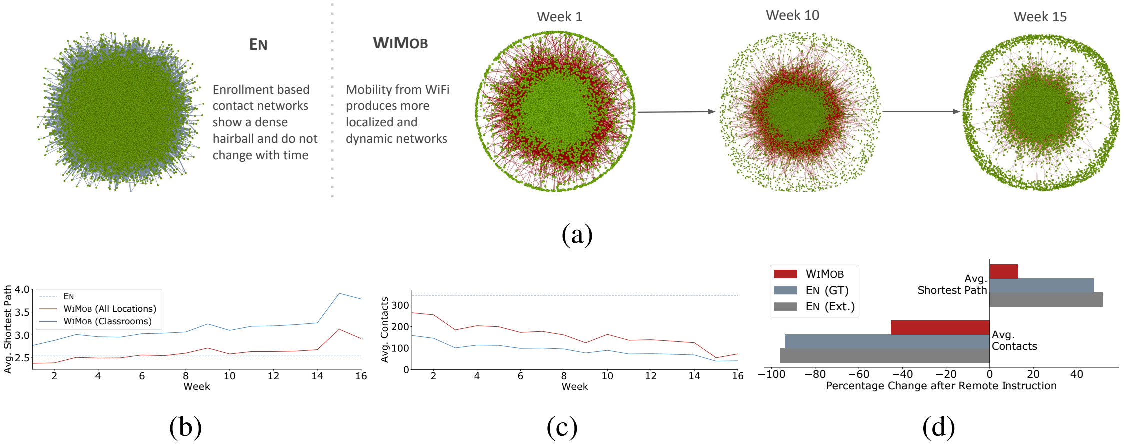


Figure 6.7: Results show difference in structural characteristics of contact networks from EN (course enrollment) and WIMOB (campus mobility). (a) In general, EN overestimates connections (grey edges) between students (green nodes) and does not anticipate changes through the semester. EN assumes 90% of students to be connected in a single component, but WIMOB reveals (red edges) that on any given week only 69% are in the largest component (those not on campus are isolated and shown in the circumference). Moreover, WIMOB reveals that density of connections changes over the semester. (b) EN depicts campus contacts to be connected closely into a "small world". WIMOB shows that contacts evolve over time. As mobility captures interactions outside classrooms I observe that for the first 6 weeks the shortest transmission path between people is shorter than what is reported by EN. (c) Enrolling into a course does not necessitate physically collocating with the class for extended periods (students can also choose to be entirely absent). WIMOB reflects this behavior and highlights a decline in average contacts over time. (d) These structural differences can help policymakers anticipate the effect of closure policies by describing how it fragments the underlying contact network. EN shows that remote instruction leads to a 94% reduction in contacts and 50% increase in transmission path length (similar to numbers reported in prior work [309], shown as EN (Ext.)). However, the estimate is significantly lower when measured using WIMOB. As a result, WIMOB emphasizes the limits of remote instruction policies and in turn motivates new policies that can be designed and evaluated with actual on-campus behavior.

6.2.2 WIMOB provides local, holistic and dynamic structural insights for contact networks on campus

EN approximates contact based on students enrolling for classes that could potentially collocate them in the same room during lectures. EN can provide structural insights on density of connections and disease transmission paths to inform modeling disease simulations [294]. However, such static data can overestimate attendance and ignore overlap between courses (via instructors) and organic interactions outside classes (e.g., waiting areas, dining, parties, and extra-curricular activities). Therefore, using EN can overemphasize the disease-mitigating structural changes to the network by RI interventions. By contrast, WIMOB is more grounded in community behavior as it captures multiple scheduled and serendipitous contact situations dynamically over the semester. I compared the features of contact networks constructed with WIMOB, against networks constructed with EN using data from GT for Fall semester of 2019 (August 19 – December 14), prior to any COVID-

19 reported cases in the U.S. I found that WIMOB rendered new insight into contact on campus that was invisible to the EN methodology.

WIMOB *characterizes temporal variation in proximity*

Variation in contact over the semester would naturally impact the severity of disease spread. However, EN describes a static network that does not capture such dynamics (Figure 6.7a). Instead, I found that WIMOB shows contacts got sparser over the semester. Figure 6.7c presents a notable decline in contacts after the first two weeks, which coincides with multiple orientation seminars and the so-called "course shopping" period of Fall 2019. In fact, contact decreased considerably in classrooms, with a steeper slope possibly because of reduction in attendance. WIMOB was able to reveal other observable changes, such as drop in contacts during exam period (week 15) and increase after fall recess (week 10). EN rendered a highly connected static network, which can miscalculate the speed at which a disease spreads. By contrast, the longitudinal behavior represented by WIMOB can help universities anticipate disease spread more accurately.

EN *overestimates contact-based risk*

Campuses can assess risk of an outbreak by characterizing the number of individuals that would be at risk of infection through contact. In our study, EN indicated 99% of the individuals on campus were clustered in a single component — if any of them would have been infected in Fall 2019, the entire component would be at risk. From the lens of EN a virus can exhaust an entire population with infection very early. However, WIMOB showed that only 69% of the population was connected in a single component. This difference is because WIMOB can distinguish how many individuals are active on campus. Therefore, WIMOB provides a pragmatic estimate of risk by grounding it in local occupancy and helps campuses budget for resources better.

WiMob *reveals different paths for disease transmission*

Reports suggest that a key contributor to cases in the pandemic is actually clustering of individuals in non-academic spaces [280]. However, En does not depict a holistic view of campus contact. It is limited to classrooms and, therefore, fixates on contacts in lectures, while ignoring other spaces. In fact, WiMob showed that in the first 6 weeks of Fall 2019, the shortest path among individuals was smaller than that approximated by En (Figure 6.7b). WiMob revealed new paths in the contact network from situations outside classes. On a given week, WiMob showed the average shortest path with contact is $3.26(\pm 0.5)$ when only considering lectures, whereas capturing all contexts reduced the average shortest path to $2.67(\pm 0.28)$. Characterizing shorter pathways is crucial for policymakers as closure policies by design aim to disconnect these pathways.

En *overemphasizes the impact of remote instruction*

Prior work used En to posit that RI reduces contact and in turn significantly fragments the network for disease spread in universities [309, 310]. To compare policy effectiveness with WiMob, I operationalized RI in this study:

Remote Instruction (RI): The status quo for data-driven policies offers strictly online instruction for large class enrollment, while continuing the other classes in person. When using En to model contacts, I implemented RI by removing connections between students who were only in contact through courses of size ≥ 30. When using WiMob to model contacts, I removed connections between students if they were only connected because of collocations during scheduled lectures of such courses.

I evaluated the effectiveness of such a policy if it were applied in Fall 2019, with both WiMob and En. Figure 6.7d shows that RI with En reduced contact by 94% and increases shortest path by 50%. However, the same intervention with WiMob showed a relatively milder impact (contact reduction 45%; shortest path increase 11%). This reinforces that contact outside courses were significant and remain unaffected by enrollment-oriented policies like RI. Instead, WIMOB provided a more encompassing view of the structural effects to a network and motivates design of more impactful closure policies.

6.2.3 Epidemiological model built with WIMOB shows that LC yields better infection reduction outcomes with lower burden

As outlined above, EN does not comprehensively capture the contact on campus. By contrast, contact networks built with WIMOB demonstrate new structural insights, which are critical to describe disease spread. A campus is composed of many different spaces, and EN does not have the flexibility to design closure of such spaces or assess its impact. These drawbacks naturally motivate a new approach to design interventions. Since WIMOB mitigates the limitations of EN, I leveraged it to demonstrate the effectiveness of localized closure (LC) policies.

WIMOB *can model RI and LC interventions with various configurations*

Prior works show a few locations are responsible for majority spread [288] and restricting movement between them leads to greater control [311]. In addition to RI, I modeled LC, which I formalize as follows:

Localized Closure (LC): I identified rooms–level spaces that are highly central location nodes in the network. I removed contacts between people who are only connected because of collocating at these locations. While, I employed various centrality algorithms to identify such locations, for the results discussed in this section I use *PageRank* [295]).

I found that, if COVID-19 spread through a regular semester, the cases rose after 7 days (Figure 6.8a). Therefore, I applied both RI and LC interventions after the first week.

To make the comparisons between the closure policies, I established fixed budgets to design LC based on the resource utilization on RI. I considered 2 kinds of budgets, (i)

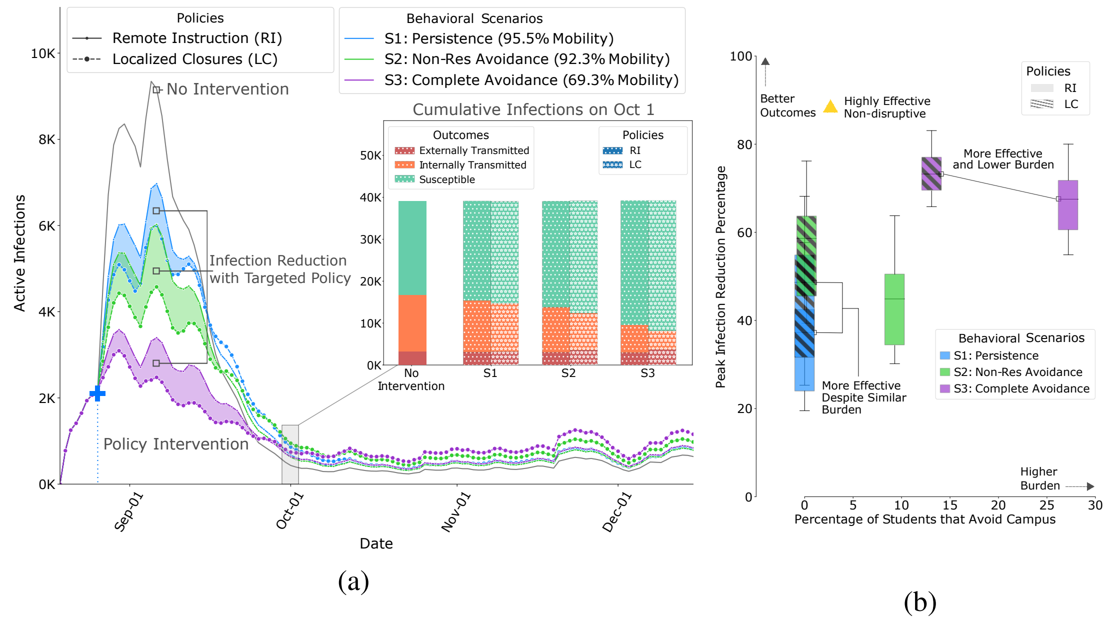


Figure 6.8: Results of policy interventions with our calibrated ABM on contact networks from Fall 2019, derived from WiMob (a) This graph compares the mean active infections between LC and RI. LC show improved outcomes (shaded regions) even when constrained to the same restrictions of RI policies. (a)–*inset:* After the first wave, even though LC shows slightly higher active infections, the cumulative infections are still lower, especially those that are a result of internal transmission on campus. (b) Outcomes of policies within the same behavioral scenario are shown with boxes of the same color (RI policies are solid, LC policies are hatched) and box heights represent the 2.5^{th} and 97.5^{th} percentile. In *S1*, even though LC and RI are equally burdensome in terms of students avoiding campus, LC shows improved outcome on peak reductions. In fact, for the other scenarios, LC shows better outcomes than RI, without forcing as many students into online schedules, and, therefore, being even less burdensome with greater impact.

mobility reduction — to depict space use on campus, and (ii) risk of exposure — to reflect testing capacity. Also note, response to closure policies can lead to unpredictable side-effects in campus behavior, particularly when a student's schedule is entirely online. Therefore, I design policies within three behavioral scenarios (each with a varying budget):

S1: Persistence: Irrespective of the locations closed or classes restricted, individuals continue their other visiting behaviors.

S2: Non-Residential Avoidance: Non-residential students stop all visits to campus if they enrolled in at least 3 courses and the policy forces their entire academic schedule online.

S3: Complete Avoidance: Same as S2, but even residential students avoid campus based on their schedule.

Similar to other works that model closure [309, 312], I assume that when a location is shutdown, the individuals who ought to have visited that location isolated during the time.

To devise interventions, WiMob estimated how RI uses the budget and then designed LC to match this budget under every behavioral scenario Table 6.9 describes how the budget for each policy varies.

I present differences between LC and RI based on three infection reduction outcomes; peak infections (maximum active cases on a given day), internal transmission (exposure from infected individuals on campus), and total infections (cumulative cases at the end of the semester). Additionally, I measured the burden of policy interventions with the number of locations closed — requires resources to monitor and maintain super-spreader locations, the percentage of students that avoid campus — disruption to learning outcomes [71, 73], and the percentage of individuals completely isolated — worsens mental wellbeing [313].

LC cause greater reduction in peak infections, while affecting fewer locations

Controlling peak infections relaxes the burden on a university to support positive cases for any given day, and allows resources to be distributed over time. In all behavioral scenarios of our simulation of Fall 2019, I observed that the peak reduction was significantly better in LC (Figure 6.8) than RI. While RI impacted 58 different locations (classrooms and lecture halls), in *S1* and *S2*, LC achieved better outcomes by closing fewer locations. For example, in *S2*, RI achieved a 28.9% peak reduction, but LC showed reductions of 49.3% (mobility budget) and 48.1% (exposure risk budget). This was attained by closing 38 or 50 locations respectively. Therefore, with such policies, policymakers need to restrict fewer locations to remarkably minimize the pressure of active infections on campus (e.g., diagnoses, treatment, quarantining).

Table 6.9: Comparison of policies in terms of controlling the disease and impacts on campus in Fall 2019.

Behavioral Scenario	S1: Persistence			S2: Non-Res Avoidance			S3: Complete Avoidance		
Policy	RI	LC		RI	LC		RI	LC	
Budget	-	Mobility (95.5%)	Exposure Risk (18800)	-	Mobility (92.3%)	Exposure Risk (16900)	-	Mobility (69.2%)	Exposure Risk (12700)
Infection Reduction Outcomes									
Peak Infections (%)	25.34	36.92**	34.30**	35.44	49.33**	52.19**	61.62	69.34**	64.44**
Total Infections (%)	6.99	10.63**	8.19**	14.88	13.96*	15.67	33.00	33.4	26.94**
Internal Transmissions (%)	17.13	22.62**	21.01**	27.58	35.35**	39.20**	54.00	70.89**	60.90**
Burdens on Campus									
Locations Affected	58	18	19	58	38	50	58	192	124
Students Avoiding (%)	0	0	0	9.30	0.20	0.45	27.21	12.45	6.57
Completely Isolated on Campus (%)	5.42	8.40	8.40	5.95	5.72	5.71	7.09	5.18	5.23

Within each behavioral scenario, I performed the Kruskal-Wallis H-Test [314] to compare outcomes of LC with RI. I found that LC leads to significantly improved peak infection reduction and internal transmission. In terms of reduction in total infections, the outcomes were comparable in general but varied by specific scenarios. In addition, every policy also exerted some burden on campus, either in terms of locations affected, students avoiding campus or isolation. I observed that LC policies focus on fewer locations (except in *S3*). Moreover, these policies affected fewer student's schedules and therefore fewer people avoid campus due to completely remote schedules. Finally, LC does not increase the percentage of people completely isolated on campus (p-value: < 0.01:*, < 0.001:**).

LC lead to comparable reduction in total infections, while keeping more students on campus

Universities want to minimize the number of infected cases while ensuring majority of the population remains active on campus to continue successful operation. In Scenario *S1*, the total number of infections reduced by both LC was more than the reduction shown by RI. For other behavioral scenarios the total infection reduction between policies was similar. In contrast, the impact the policies had on the student schedules was remarkably different. RI forced multiple students to adapt to fully online schedules. In Scenario *S2*, 9% of students did not visit campus and in *S3*, 27% of students did not visit campus. On the other hand, in LC, the number of students expected to avoid campus could be as low as 0 and never exceeded 12%. Besides sustaining economic loss to the campus, remote instruction can increase anxiety among students and hinder learning outcomes [73, 74]. Compared to RI, LC offers policymakers a way to defend against turnover in the student population, without compromising overall control of disease spread (Table 6.9). Limiting the number of students that avoid campus helps preserve on-campus businesses [315, 316] and minimally disrupts the student wellbeing.

LC cause greater reduction in internal transmission without causing further isolation on campus

Universities want to limit infection spread, but they must also ensure that aggressive policies do not worsen mental wellbeing of the community. In terms of internal transmission the reduction was significantly larger with LC (Table 6.9). However, when LC restricted the infections early in Fall 2019, it left more individuals susceptible to external transmission. College student behavior outside campus on weekends and breaks is known to impact local transmission [317]. When policymakers consider LC they should also consider policies on re-entry or required testing based on off-campus activities. In terms of isolating individuals on campus, it's notable that LC and RI were similar in *S2*. Interestingly, in *S3*, where

LC closed more than 100 locations, the percentage of isolated individuals per week was less than that of RI. This finding implies that LC can keep individuals on campus without forcing them into complete isolation. Therefore, LC can help alleviate concerns of closure interventions that increase loneliness and limit social connectedness [318].

LC identifies a wider variety of auxiliary spaces.

By using WIMOB to design LC I was able to identify locations for closure at the granularity level of rooms, including unbound spaces such as lobbies and work areas. As policy design budgets changed with every behavioral scenario I found that LC identified different types of locations for closure. First, in *S1*, I found that most locations that LC targeted are a subset of the auditoriums–like rooms where large classes would take place in Fall 2019. Note, LC needs to restrict only a few such spaces to utilize the same budget as RI. This was because, under *S1*, RI policies only altered visits to lectures, while these spaces are used for other purposes during other times (e.g., club activities and seminars). I also found that LC targeted 'high traffic' locations like conference center lobbies which are typically used as waiting areas or for networking events. Next, in Scenario *S2*, I saw that in addition to spaces mentioned earlier, interestingly LC further restricted the use of smaller rooms (occupancy $13 - 35$) which would not be affected by RI (as only classes of size ≥ 30 are offered online). LC also targeted areas in the recreation center (which includes locker rooms and indoor courts for $4 - 20$ people). This insight indicates that WIMOB accounts for a diverse set of student activities. Lastly, in Scenario *S3*, LC targeted closure of activity in far more spaces than RI. However, the better outcomes can be attributed to the fact that LC diversified the potential restriction areas. LC restricted heavily used small study rooms or breakout rooms (for $1 - 6$ people). Furthermore, it restricts use of spaces where multiple small groups of people can organically assemble, such as cafes, dining halls, and reading areas. I also observed that LC restricted activity in about 10 Greek Houses but does not target other housing areas — demonstrating its ability to restrict social behavior that could amplify disease spread.

Sensitivity and robustness analyses

So far, I used an ABM calibrated on the positivity rate of the first 5 weeks of Fall 2020. This rate can be influenced by many latent factors (e.g., mask-wearing, hand washing, distancing, and compliance) on a campus. To study any effect of these variations, I also calibrated on different time windows throughout the semester. I calibrate on weeks $5-9$ and $10-14$ in Fall 2020, and validate on the remaining semester. In both cases, compared to RI, I found that LC still exhibits better reduction in peak infections (up to 90%) and internal transmission (up to 77%). In the original calibration, LC maintained the same level of total infections as RI, but with the new periods I found total infections were substantially less than RI. Another important variable for positivity is the wider context of the campus e.g. urban/rural, the surrounding county, city, etc. To investigate this, I also calibrated our ABM on the positivity rate of different universities in the US in Fall 2020 (along with information from their county to seed external cases). Consider this as a hypothetical where the mobility of the GT community remains the same but disease outcomes resemble a different campus. I calibrated on data from University of Illinois at Urbana-Champaign and University of California, Berkeley. I found no remarkable differences from our findings with GT.

6.2.4 Summary

NPIs are the first line of defense for campuses to respond to contagious diseases like COVID-19 [319, 320]. Traditional, approaches such as EN can misconstrue contact on campus, leading to policies like RI, which can have broad impacts despite their effects on curbing the disease spread. This study demonstrated a passive sensing framework that repurposed logs from a managed WiFi network (WIMOB) to drive organizational decisions for effective localized closure policies (LC). Overall, WIMOB presents an attractive and practical method to inform better public health policies.

CHAPTER 7
METHODOLOGICAL CHALLENGES IN EFFECTIVELY MODELING PASSIVELY SENSED DATA TO INFER PSYCHO-SOCIAL EXPERIENCES OF WORKERS

Over the course of the previous chapters, I have described a variety of passive sensing frameworks to improve information worker experiences. These studies demonstrated different types of everyday digital technologies that can be repurposed to describe the worker in a way that is automatic, continuous, and unobtrusive. Moreover, these frameworks can inform technologies and approaches of working by clarifying individual, team, and organizational worker-based phenomena. While these frameworks promise a variety of advantages over conventional methods like surveys, they still have their own pitfalls as they are often modelled on surveys. Therefore, if we are to deploy these frameworks, it begs the question if we are actually sensing what information workers do and feel, or just how they report it. The latter is often the basis of ground truth in the studies I have shown. And it can be challenging to determine how true the ground truth is.

RQ IV: *What are the methodological challenges of building effective passive sensing frameworks for information worker experiences?*

In this chapter I posit that passive sensing frameworks are methodologically challenged because of poor understanding of ground truth.

7.1 The Semantic Gap in Passive Inference of Mental Wellbeing: Motivation and Hypotheses

Ground truth is merely an abstraction or interpretation of what an individual is really perceiving [321]. Even for the same mental state, individuals can respond to surveys differently because of self-presentation bias [9] and non-response bias [322]. These nuances are loaded into the ground truth labels but are often ignored by passively sensed data. This lack of information, or abstraction of it, leads to a mismatch between model estimates and the actual mental state of the individual [323]. Some areas of computing refer to the abstraction between a computational model and the variable of interest as the "Semantic Gap" [324]. This gap is stark when the signals gleaned from computational data do not coincide with the factors affecting the ground truth. I argue that, for real-world longitudinal studies of mental-wellbeing, the semantic gap limits certain passive sensing models due to the nature of ground truth measures.

Consider a dominant form of mental wellbeing ground truth, self-reports [325, 326, 327, 145, 328]. While self-reports can approximate the psychometric component of stress (e.g., nervousness or apprehension), they do not reflect the physiological one (e.g., increase heart rate) [329]. A survey response may be influenced by the retrospective psychological effect at the time of reporting [330, 331]. Importantly, self-reports are sensitive to psychosocial factors such as recall bias, impression bias, and self-censorship [9, 322, 332]. An individual's self-report can be disconnected from their behavior because they are uncomfortable disclosing the severity of their state [321, 322]. These psychosocial influences on self-reports are invisible to typical approaches of passive sensing, which focus on individual physical behaviors, such as activity duration, mobility, and device usage. Despite the same bodily response to watching an intense horror movie compared to being reprimanded by a supervisor; a survey response could report different stress severity for each experience. On the other hand, physiological measures of ground truth might indicate the same severity, but ignore the negativity associated with the experience. Therefore, predictive models trained on an individual's activity data can be limited simply because of a mismatch between the choice of feature representations and the type of ground truth measurement. I believe this represents a semantic gap. In this chapter I empirically demonstrate that this gap exists in our domain and prescribe approaches to mitigate it.

Other computing areas plagued by the semantic gap teach us that this gap is narrower when the low-level representations (e.g., passively sensed features) and high-level representations (e.g., ground truth values) for the same concept are semantically analogous [323]. This fundamental informs my inquiry. For example, posts on online social media can encapsulate the same psycho-social influences that interfere with self-reports, in terms of self-disclosure [333] and censorship [334]. This raises the question, *RQ1: "Compared to behavioral signals, do social signals have a smaller semantic gap with psychological interpretations of wellbeing?*". My study investigates two specific hypotheses:

H1a. Features extracted from social media posts are more predictive of self-reported *anxiety* than features extracted from sensors of offline physical activity

H1b. Features extracted from social media posts are more predictive of self-reported *stress* than features extracted from sensors of offline physical activity

By contrast, physical behaviors from offline sensing are semantically closer to physiological aspects, such as arousal [335, 336, 337]. This raises the question, *RQ2: "Compared to social signals, do behavioral signals have a smaller semantic gap with physiological interpretations of wellbeing?*". I investigate this with a specific hypothesis:

H2. Features extracted from physical activity sensors are more predictive of *high arousal duration* than features extracted than social media posts

It is typically challenging to observe the semantic gap in practical deployments because most efforts to predict mental wellbeing, focus on limited sensor streams and a limited set of corresponding ground truth measures. To mitigate this, I investigate the gap by leveraging a unique dataset that includes a variety of ground truth measures for mental wellbeing states and a variety of passively sensed data (section 3.1).

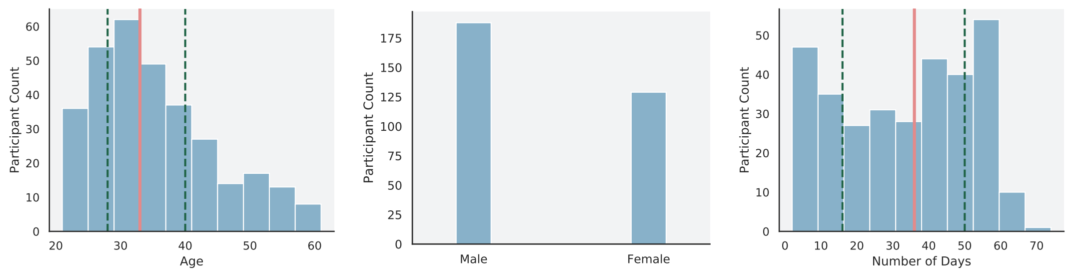
I employed the triangulation method [338] to investigate if this gap exists. I compared the predictive efficacy of different passive sensing approaches on different measures of ground truth for mental wellbeing. One relied on the offline physical activities sensed from smartphones, wearables, and Bluetooth. The other relied on the online language extracted from posts on social media. With these I built models to predict two different interpretations of a worker's mental state — the first is self-reports of state anxiety and stress, and the second is a measure of physiological arousal through a wrist-worn sensor.

Primarily, this study presents a case that characterizes the semantic gap in passive sensing for predictive wellbeing and demonstrates an approach to reduce it. By highlighting this semantic gap, my aim is neither to identify the most credible instrument of ground truth nor is it to deplore particular sensor streams. Instead, I intend to clarify why passive sensing models of mental wellbeing appear to work or fail. Moreover, acknowledging the semantic gap in our domain leads to several key implications. Through the results of my study, I encourage researchers to consciously understand the nature of ground truth labels and what factors influence that measure. And, in cases of limited sensing affordances for field study deployments, my findings motivate a more theoretical approach to sensor and modality selection for efficacious predictive studies.

7.2 Study and Data

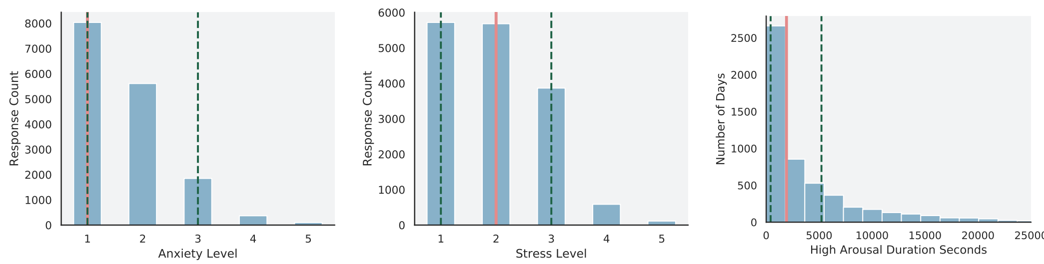
This work relied on data collected from a large multimodal sensing effort known as the Tesserae Project (section 3.1). Such a dataset is particularly appropriate for my research questions because it contains different interpretations of the ground truth (self-reported and physiological) as well as multiple sources of passive data (physical activity and social media posts). This study only considered those 317 participants that consented to data collection of offline behaviors as well as the social media collection. 129 participants reported they were female and 188 reported they were male (Figure 7.1).

Throughout the study, the phone agent facilitated Ecological Momentary Assessments (EMAs) to capture daily variations in mental wellbeing states. The self-reports for anxiety and stress formed the basis of the perceptual representation of ground truth (RQ1). Simi-



(a) Age (b) Gender (c) Study Period

Figure 7.1: Summary of participants. The solid red line indicates the median and the dotted green lines indicate the inter-quartile range.



(a) Anxiety (b) Stress (c) High-Arousal Duration

Figure 7.2: Distribution of Ground Truth. The solid red line shows the median and the dotted green lines show the inter-quartile range.

larly, the wearable provided daily estimates of the duration an individual's physiology was in a state of high arousal (RQ2) [133]. Figure 7.2 shows the distribution of the ground truth.

For purposes of passive sensing, a phone application was installed in participants' smartphones [78] and they were provided a wearable device (*Garmin Vivosmart*) along with Bluetooth beacons (*Gimbal*) [111]. These devices captured offline behaviors such as phone usage, locations, steps, sleep, and presence at home. The features for these sensors are described in this study are described in Table 3.1. Moreover, a subset of participants explicitly consented to the study of their historical and prospective social media data [94]. The corresponding linguistic features are described in Table 3.2. These different data sources represented the different comparative models analyzed in this study.

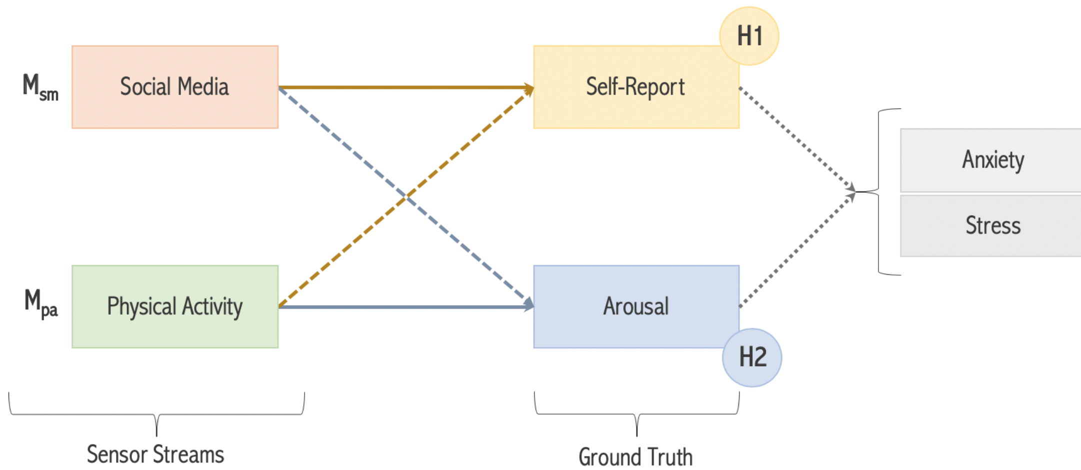


Figure 7.3: The triangulation framework helps compare different prediction models. For H1a and H1b, M_{pa} and M_{sm} predict self-reports of anxiety and stress respectively. For H2, M_{pa} and M_{sm} predict high arousal duration.

7.3 Method

To test my hypotheses, I compared the performance of two approaches for predicting different ground truth for mental states. The first used modalities with psycho-social signals, M_{sm}(social media language features), while the second is used those with behavioral signals, M_{pa}(physical activity features). It is known that neither self-report assessments nor arousal measurements alone can comprehensively capture the nuances of anxiety or stress. Therefore, I compared different approaches to predict them as a means to disentangle the semantic relationship between low-level computer representations and high-level human wellbeing constructs. This approach was motivated by previous works that used quantitative data triangulation [338] to elucidate the complexities in different phenomena. The general framework of triangulation [339] is suitable for deconstructing predictive analysis of wellbeing using passive sensing because it "adds rigor, breadth complexity, richness, and depth to any inquiry". Figure 7.3 illustrates an overview of my investigation framework.

7.3.1 Feature Engineering

This study refers to both anxiety and stress, in terms of a 'state', because these constructs are susceptible to change in short periods of time. In the scope of this work, the ground truth measures (both self-report and arousal) were collected at the day-level granularity [111]. Prior works in pervasive sensing for mental wellbeing [340, 341, 342] motivated me to analyze behavior not just during the day of ground-truth measure, but also in periods preceding it. Even theoretically, mental states are indicated by general trait behavior that changes but is less sensitive [331]. Therefore, my approach accounted for the historic sensor data to approximate the target concept. To this end, I first collated features that spanned a period prior to the prediction day.

Feature Windows

The predictive models I built for both M_{pa} and M_{sm} to consider a window of time for the features. For instance, to predict the state-anxiety (H1a) for day n the model analyzed features in a span of d days before n. Here, d dictates the fixed window size.

Physical Activity. Since offline sensors could continuously monitor individuals I varied the window size between $1 - 15$ days. This results in $f \times d$ dimensions if f is the original set of features computed for each day and d is the window size. Importantly, these sensors were not active before the first self-report was collected for any participant [111], therefore this modality was limited in how far the window could stretch retrospectively. Since the average participants had 31 labels the upper limit of 15 days was chosen to balance the remaining days for evaluation. If a window of, say, 31 days was chosen then in most cases, only the most recent label would have physical activity features for every day while all days before that would have empty data.

Online Language. Unlike physical activity data, which provided a near-continuous and contiguous signal, the online language data obtained from social media is extremely sporadic. Social media can be considered a form of "virtual sensor" that capture rich momentary events, which occur irregularly [343]. This is inherent to the approach as people do not post regularly, thus making social media platforms approximate event-based sensors. Thus, the window size for this modality varied between $30 - 180$ days, with a shift of 30 days between each window. In contrast to offline sensors that were only instrumented after enrollment, social media allowed me to access data prior to enrollment and could, therefore, support a much broader window [94].

Prepossessing

This section elaborates on my methodology for imputing missing values and standardizing features in windows.

Physical Activity. On certain days particular features could be missing due to participant compliance (e.g., the participant did not charge a device or data failed to log). Consequently, I imputed the missing values of a feature by substituting it with the mean of that feature for an individual for a given window. To demonstrate, if a feature value was missing for an original feature f^a on day d_j, then the average will be $\sum_{i=1}^{d} f_i^a/d$, where d is all days the feature was not null. After this the features were standardized by subtracting the mean of the feature values and dividing it by the standard deviation. Similar to the imputation, the standardization procedure was also applied within windows, i.e., the average and standard deviation for any feature f_i^a, was calculated on $[f_1^a, f_d^a]$ where d is the window size.

Online Language. Empty values occur much more frequently because most participants did not post everyday. Because of this limitation, filling in missing values with averages could lead to washing out any true variations. Therefore, I heuristically rejected windows that had fewer than 1 post per week on account of low density. This was followed up by the approach described earlier where both imputation and standardization are applied within windows.

7.3.2 Feature Processing and Model Training

I developed different non-linear regression models for M_{pa} and M_{sm} to estimate self-reported state anxiety (H1a), perceived stress (H1b), and objectively measured high-arousal duration (H2). In particular, I trained models with both modalities using estimators that rely on ensemble learning because these approaches "reduce the variance — thereby improving the accuracy" of estimates [344, p. 1]. The *Random Forest* regressor aggregates independent decision trees, each of which learns on a random sample of input features [345]. *Gradient Boost* learns incrementally over time by increasing the importance of poorly estimated observations in every subsequent iteration [346]. An additional variation to this is Extreme Gradient Boosting (*XGBoost*), which is both robust to noise and designed to deal with sparse input features [347], such as those extracted from social media data. Moreover, a different model was built for every window size and each model was trained using a 5-fold cross-validation method. Additionally, the grid search approach tuned the parameters for each model [273]. Since the information used to predict the target value for each day was spread across a window of d days, it leads to $f \times d$ dimensions, which can sabotage the training because of the *curse of dimensionality*. To tackle this, I employed certain feature transformation and reduction techniques to improve the model training. These processing approaches are applied to each model separately, i.e., it is unique to the window size. Given the cross-validation approach, these feature processing steps were "fit" only to the training data without incorporating any of the observations in the testing folds.

Coefficient of Variance

First, I estimated the variance explained by each dimension measuring the *coefficient of variance* (CV) [146]. With a conservative bound, I remove dimensions that are beyond 1 standard deviation of the average CV. For the linguistic features included in the M_{sm} models, this typically led to a dimension reduction by $20 - 26\%$ with windows varying between $30 - 180$ days. For the physical activity features in M_{pa}, this led to a reduction of

$32 - 14\%$ for windows of size $1 - 14$ days. *Note: The M_{pa} model used in H2 does not use this aspect of the pipeline because it produces a better model without this selection.*

Principal Component Analysis

Next, I further reduced the dimensions by performing PCA on the remaining dimensions [348]. This approach identifies latent components in the data (linear combinations of existing dimensions) that explain maximum variance. The first set of principal components that can cumulatively explain more than 90% of the variance in the data were selected as dimensions going forward. For M_{pa}, between window sizes of $1 - 14$ days, this process reduced dimensions by $62 - 84\%$ respectively. Similarly, for M_{sm}, between window sizes of $30 - 180$ days I found a reduction of $51 - 86\%$.

Mutual Information

Lastly, for M_{sm} I included a final shortlist of dimensions based on mutual information between the input dimensions and the target variable [349]. Based on the mutual information scores, this process selected the top 10 percentile dimensions. It is important to note that this procedure is both unnecessary and detrimental to apply on the features of M_{pa} as these models had lower dimensionality to begin with and reduction beyond the PCA described earlier generated weaker models.

7.4 Results

I study the semantic gap by comparing different prediction approaches with an analytic process grounded in the data triangulation framework [338]. This framework enabled me to methodologically evaluate heterogeneous approaches to understand the same phenomenon [339]. The approaches I compared in this study differ in terms of both data source and methodology. Therefore, for each , this study addresses the research questions on the basis of the best models for M_{pa} and M_{sm}.

Table 7.3: Summary of between–models comparison for self-reported anxiety
(‘-’:$p < 1$, ‘.’:$p < 0.1$, ‘*’:$p < 0.05$, ‘**’:$p < 0.01$, ‘***’:$p < 0.001$)

	Pearson's R		SMAPE	
Regressor	M_{pa}	M_{sm}	M_{pa}	M_{sm}
Random Forest	0.34***	0.56***	0.18	0.16
Gradient Boost	0.27***	0.51***	0.19	0.17
XGBoost	0.27***	0.51***	0.19	0.17
Window Size (days)	13	30	13	30

Within Modality Comparisons. The best model was chosen based on the highest pooled *Pearson's* correlation between the true values and the predicted values. Specifically, I pooled together the predictions from each cross-validation fold and then computed the correlation with the ground truth. This approach is robust to heterogeneity in target variables' distribution between folds and provides a more generic measure of performance [350]. I used the *Pearson's* correlation coefficient because it spans all samples to describe a complete relationship, is not sensitive to the distribution of samples, and does not assume normality [209]. This correlation contrasts a model's input features and the target variable. For internal validity of the regression models, I compared the *Symmetric Mean Absolute Percentage Error* (SMAPE) against an arbitrary regression model that always predicted the mean of the training data.

Between Modality Comparisons. Once the best models of M_{pa} and M_{sm} were identified I validated comparisons between M_{pa} and M_{sm} by performing a permutation test [349, 274]. Essentially, I attempted to reject the null hypothesis that a random set of features in a similar feature space (range and dimensionality) will still perform better than the worse model [146]. As a result, I permuted random features in the same space and compute the probability ($p-$value) of such an arbitrary model improving over the benchmark.

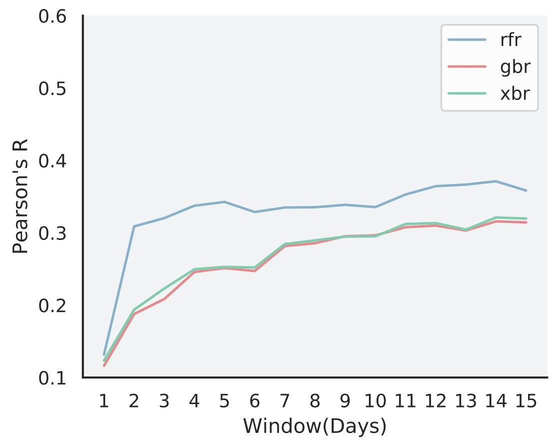
(a) M_{pa} **(b)** M_{sm}

Figure 7.4: Comparing models with different window length to predict anxiety

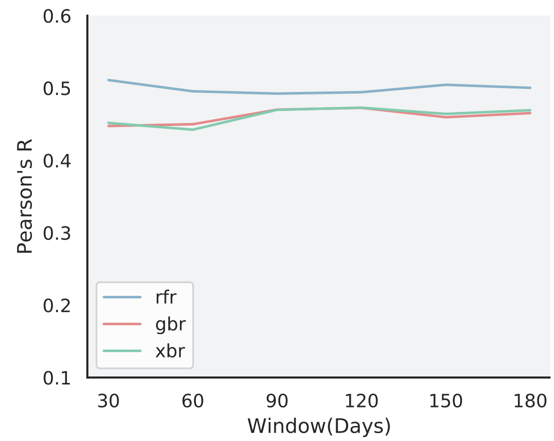
7.4.1 RQ1: Semantic Gap in Predicting Psychological Aspects of Wellbeing

H1a: M_{sm} *is a better predictor of self-reported anxiety*

We find language on social media to be more indicative of self-reported state anxiety when compared with physical activity from offline sensors. With physical activity features, we find the best model for M_{pa} to be with a window length of $d = 14$ and using the Random Forest regressor (Figure 7.4a). This model recorded a *Pearson's r*$= 0.34$. In comparison to an arbitrary regressor, which demonstrated a *SMAPE*$= 0.20$, this model shows a *SMAPE*$= 0.18$, a 10% improvement over the baseline. By contrast, for the same target variable, the best M_{sm} model was at $d = 30$ with a Random Forest regressor (Figure 7.4b), which yields a *Pearson's r*$= 0.56$. In comparison to the baseline (*SMAPE*$= 0.21$), this model improves by 30% (*SMAPE*$= 0.14$). Between models, we see the *Pearson's r* in the anxiety values predicted by M_{sm} to be 64% better than values predicted by M_{pa}. To test the robustness of this comparison I ran the pipeline for M_{sm} 1000 times with randomly generated permutations of the feature values and find the probability of improvement over M_{pa} to be less than 0.001. As a result, this asserts M_{sm} is more predictive of self-reported state anxiety than M_{pa}, and this supports hypothesis H1a (Table 7.3).



(a) M_{pa}



(b) M_{sm}

Figure 7.5: Comparing models with different window length to predict stress

Table 7.4: Summary of between–models comparison for self-reported stress
('-':$p < 1$, '.':$p < 0.1$, '*':$p < 0.05$, '**':$p < 0.01$, '***':$p < 0.001$)

	Pearson's R		SMAPE	
Regressor	M_{pa}	M_{sm}	M_{pa}	M_{sm}
Random Forest	0.37***	0.51***	0.18	0.17
Gradient Boost	0.31***	0.44***	0.18	0.18
XGBoost	0.32***	0.45***	0.18	0.18
Window (days)	14	30	14	30

H1b: M_{sm} *is a better predictor of self-reported stress*

Similar to the previous result, language on social media is more predictive of self-reported stress than physical activity from offline sensors. In the case of M_{pa}, the window length of $d = 14$ with the Random Forest regressor (Figure 7.5a) emerged at the best model with a *Pearson's r*$= 0.36$. This improved on the baseline (*SMAPE*$= 0.20$) by 10% (*SMAPE*$=$ 0.18). On the other hand, the best M_{sm} model was at a $d = 30$, also with Random Forest shows a *Pearson's r*$= 0.51$ (Figure 7.5b). Compared to the baseline (*SMAPE*$= 0.20$), this model had a *SMAPE*$= 0.17$, i.e., a 15% improvement. When comparing the two models, we find the *Pearson's r* of M_{sm} to be 37% better than that of M_{pa}. The permutation test

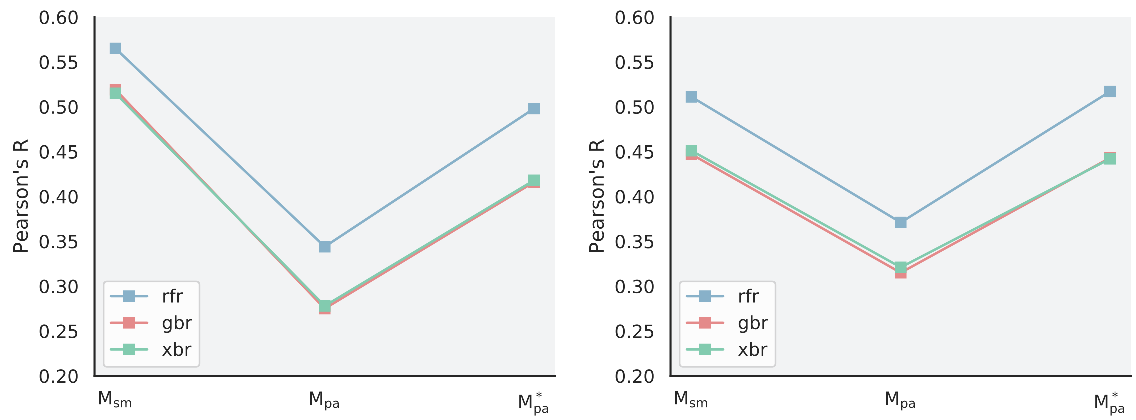
Table 7.5: Social features extracted from offline sensors

Category	Features	Stream
Colocation	Time of first and last interaction, number of interactions, number of unique participants, duration of interactions, percentage alone, percentage with at least one /two /three others	

was run 1000 times for random versions of M_{sm} and improved over M_{pa} less than 0.001 of the time. Based on the results, M_{sm} was a better predictor of self-reported stress than M_{pa}, and therefore the hypothesis H1b holds (Table 7.4).

Post-Hoc Analysis

The results of the experiments argue that features extracted from social media posts can encapsulate analogous phenomena and therefore predict the target variable better. However, social signals can be derived from data acquired through offline signals as well. Since offline interactions are subject to similar presentation effects [351], we performed an additional experiment that augments M_{pa} with some physically sensed social features. In particular, we used the Bluetooth beacons to identify social behaviors, such as the time of first interaction, number of unique interactions, and their duration (Table 7.5). I included these features in the models used to test M_{pa} to predict the ground truth. The chapter refers to this combined modality as, M^*_{pa}. In fact, the pipeline used for M_{pa} is the best framework for M^*_{pa} as well. For anxiety, the optimal results were produced with a random forest regressor at a window length of $d = 13$ where the Pearson's r is 0.49. Albeit still less than M_{sm} (Pearson's r is 0.56), this was markedly more than the best model for M_{pa} (Pearson's r is 0.34) by 64%. Actually, for predicting stress, the best results emerged with the same regressor and same window length (Pearson's r $= 0.51$). Not only was it better than M_{pa}(Pearson's r $= 0.37$) by 41%, it was comparable to M_{sm} as well (Pearson's r $= 0.51$).



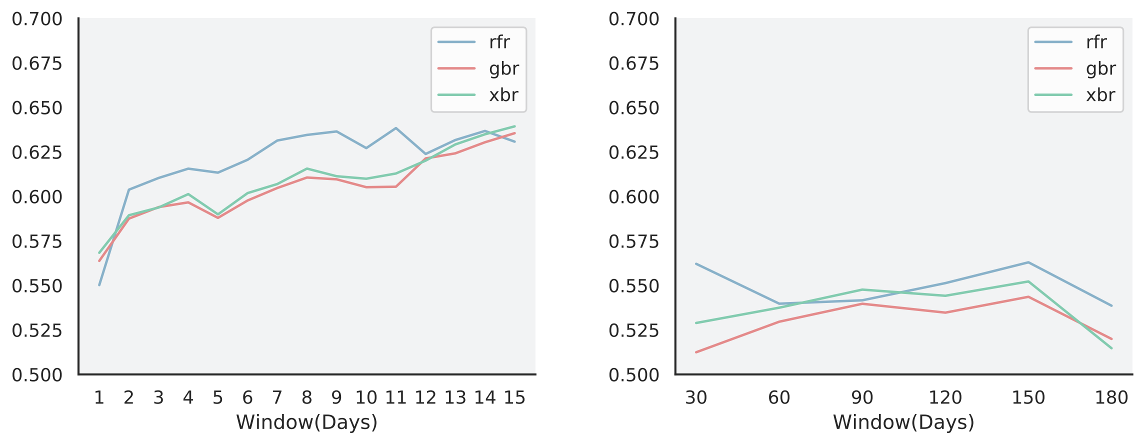
(a) Anxiety

(b) Stress

Figure 7.6: Comparison between best models of different modalities (RQ1)

Interpretation

The results for predicting self-reported anxiety and self-reported stress support my hypotheses towards my first research question, which investigates if features encapsulating social signals reduce the gap with self-reported measures of wellbeing (Table 7.6a and Table 7.6b). Mental wellbeing constructs like anxiety and stress have different aspects. Self-reports are skewed to capture the psychological aspects [330] that might not be concordant with how the individual actually behaves. Moreover, self-reports are influenced by many social effects of self-presentation such as impression management [9, 352] and response bias [322]. These effects are inconspicuous to sensors that capture physical activity, even though that data can be modeled to predict such self-report (evident from the improvement on the baseline). By contrast, data sourced from social media are weaved with similar ecological effects that could influence an individual's self-report. For example, self-disclosure [333] and self-censorship [334] are both factors that affect the language posted online. This could explain why M_{sm} exhibited better results than M_{pa} when trying to predict self-reports. Relatedly, incorporating more explicitly social features in offline sensing also shows an improvement in the prediction (M^*_{pa}).



(a) M_{pa} **(b)** M_{sm}

Figure 7.7: Comparing models with different window length to predict high-arousal duration

Table 7.6: Summary of model comparison between–models to predict high-arousal duration
('-':$p < 1$, '.':$p < 0.1$, '*':$p < 0.05$, '**':$p < 0.01$, '***':$p < 0.001$)

	Pearson's R		SMAPE	
Regressor	M_{pa}	M_{sm}	M_{pa}	M_{sm}
Random Forest	0.63***	0.56***	0.45	0.46
Gradient Boost	0.60***	0.54***	0.46	0.46
XGBoost	0.61***	0.55***	0.43	0.46
Window (days)	11	150	11	150

7.4.2 RQ2: Semantic Gap in Predicting Physiological Aspects of Wellbeing

H2: M_{pa} is a better predictor of objectively-measured high-arousal duration

While predicting high-arousal duration the model built with physical activity features was better than the corresponding model built with social media language features. We find the best model for M_{pa} to be at window length of $d = 15$ with the XGBoost regressor (Figure 7.7a), which showed a *Pearson's r*$= 0.63$. This surpassed the baseline (*SMAPE*$=$ 0.54) by 16% (*SMAPE*$= 0.45$). In comparison, the best performing M_{sm} model occured at

$d = 150$, also with Random Forest, which yielded a *Pearson's r*$= 0.56$ (Figure 7.7b). This model had a *SMAPE*$= 0.43$, i.e., a 20% improvement on the baseline (*SMAPE*$= 0.54$). In comparison to M_{sm} the *Pearson's r* of M_{pa} is 13% better. To reject the possibility of chance improvement, from 1000 randomly generated permutations of M_{pa} less than 0.01 feature sets improved over M_{sm}. These results indicate that M_{pa} is a better predictor of self-reported stress than M_{sm} and therefore supports hypothesis H2 (Table 7.6).

Post-Hoc

Similar to the analysis performed in subsubsection 7.4.1, I further experiment on predicting physiological wellbeing by including offline sensed social features (Table 7.5). The argument to pursue such an analysis in the light of RQ1 was to estimate the effects of social signals from alternative sources to reduce the potential semantic gap. However, in RQ2 testing a prediction with M^*_{pa} is to explore how social factors interact with physical signals to predict physiological aspects of wellbeing. On experimenting with M^*_{pa} we find that a random forest regressor at a window length of $d = 13$ yielded the best result of *Pearson's r* $= 0.69$. Compared to the large boost we observed in predicting self-reports, adding social signals to predict objective measurements only augmented M_{pa} (Pearson's r is 0.63) by only 9%. While this is still noticeable, I believe the improvement is limited by the nature of the additional signal (psycho-social) in comparison to the representation that is being predicted (physiological). Therefore, although additional features can lead to some increment in performance, large boosts can be achieved when a model is augmented by semantically similar features (subsubsection 7.4.1).

Interpretation

The findings for predicting high arousal duration support my hypotheses towards the second research question, which speculates a reduction of the semantic gap in predicting objective measures of wellbeing by modeling features with behavioral signals (Figure 7.8).

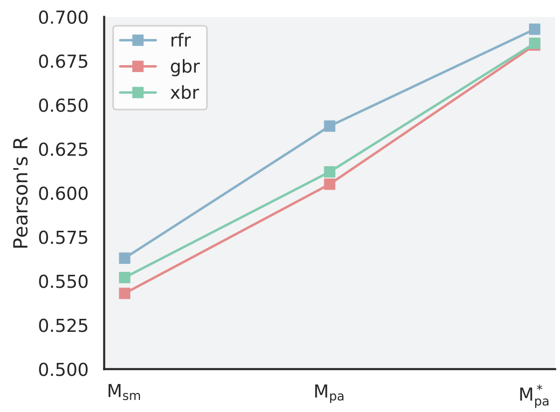


Figure 7.8: Comparison between best models of different modalities (RQ2)

This question was proposed to provide divergent validity to the first question and reinforce the quantitative data triangulation method of validation [338]. The physiological aspects of wellbeing can often be independent of what individuals report [330]. It can be subject to inherent beliefs, other subjective factors, and confounding mental phenomena [353]. On the other hand, the physiological experience of the individual remains consistent. Furthermore, the physical behaviors of an individual are coupled with physiological responses to wellbeing constructs like stress. For example, increased activity can reduce arousal by expending energy [336], while reduced sleep can be the result of increased arousal [337]. This kind of information is challenging for passive sensing through online traces to perceive as people only present a part of their selves on such platforms. Accordingly, M_{pa} performed better in this regard due because offline modalities that continuously capture an individual's functioning can illustrate richer representations of their behavior.

7.4.3 Participant-Independent Models

The models described in subsection 7.4.1 and subsection 7.4.2 were validated by using some observations in the training folds while others are used in testing folds (each participant had an average of 33 days of data). Such approaches, known as "mixed-model" or "personalized-model", account for individualized routine and trait-like propensities to predict the target variable. These have been used in prior works in longitudinal sensing with

Table 7.7: Summary of best models for participant independent models.
('-':$p < 1$, '.':$p < 0.1$, '*':$p < 0.05$, '**':$p < 0.01$, '***':$p < 0.001$)

		5-Fold CV			LOPO CV		
	Ground Truth Measure	M_{pa}	M_{sm}	M^*_{pa}	M_{pa}	M_{sm}	M^*_{pa}
H1a	Self-Reported Anxiety	0.02-	0.15***	0.02-	0.02-	0.08***	0.02-
H1b	Self-Reported Stress	0.02-	0.08**	0.02-	0.02-	0.07**	0.02-
H2	High Arousal Duration	0.52***	0.03-	0.55***	0.52***	0.03-	0.55***

smartphone data [354, 340] and social media data [355]. An alternative approach to modeling sensor data is with participant-independent models which treat testing data as entirely unseen participants. These are expected to generalize better to new participant data. To inquire my hypotheses with this approach I first performed a participant-independent 5-fold cross-validation. I followed the same feature processing described in subsection 7.3.2, with the only difference being that M_{sm} performed better without any additional mutual information based feature selection (subsubsection 7.3.2). For H1, I found that the best model for M_{sm} significantly estimated both self-reported anxiety (*Pearson's r* $= 0.15$ with XGBoost when $d = 180$) and stress (*Pearson's r*$= 0.08$ with Random Forest when $d = 90$). In contrast, however, M_{pa} models did not significantly estimate the ground truth at all (*Pearson's r*$= 0.02$ for anxiety and *Pearson's r*$= 0.02$ for stress). For H2, the best model for M_{pa} significantly estimated high arousal duration (*Pearson's r*$= 0.52$ with XGBoost when $d = 1$), whereas the best performance for M_{sm} did not show significant correlation (*Pearson's r*$= 0.03$). Further, I also performed a leave-one-participant-out validation for both hypotheses. Again, for H1, we found that only M_{sm} could significantly estimate self-reported anxiety (*Pearson's r*$= 0.08$ with XGBoost when $d = 30$) and stress (*Pearson's r*$= 0.07$ with XGBoost when $d = 30$). Similarly, for H2, only M_{pa} significantly estimated high arousal duration (*Pearson's r*$= 0.52$ with Gradient Boost when $d = 1$). Even though the performance of person-independent models was lower than the personalized ones (as shown in similar studies [354, 340]), I still found that these models demonstrated a persistent semantic gap. Table 7.7 summarizes the comparison between M_{pa} and M_{sm} for the different hypotheses.

7.5 Guidelines

The findings of this study reveal the presence of the semantic gap in studies to infer mental wellbeing. I propose a set of guidelines for researchers in social and ubiquitous computing, who plan to conduct passive sensing studies to infer mental wellbeing.

7.5.1 Ground Truth Matters

> In studies of mental wellbeing, even with validated instruments to assess ground truth, researchers need to consider how this ground truth represents or abstracts the underlying mental wellbeing construct (psychological or physiological).
>
> Consequently, researchers must account for the different factors that affect these representations (e.g., social biases or behavioral artifacts) to get optimal results.

My results show that the presence of a semantic gap reflects a mismatch between what computational models represent and what different ground truth represent for the same mental wellbeing state. From the perspective of computer scientists, the ground truth is considered an unquestionable "gold standard". The literature has discussed several challenges to passive sensing [356], such as choice of device, application, duration, and sampling rate. My findings extend this list with a focus on ground truth representations. This study demonstrates a case that urges conscious consideration of the ground truth's sensitivity to ecological factors. Many studies in the community tend to acquire ground truth *in situ* [357, 79, 78, 111] but it distances the researchers from carefully observing the circumstances of ground truth measurement. In reference to the uncertainty of ground truth labels, Plötz's third postulate for machine-learning on sensor data states, "there is no ground truth" [358]. I situate this in the context of passively inferring mental wellbeing. Do participants respond to anxiety questions immediately after stressful incidents or do they summarize the experience of their day? Do they actually report how they felt or are their responses describing the state they wanted to be in? These concerns are not only challenging to quantify but also opaque to researchers and sensors [357]. **However, acknowledging the semantic gap can help researchers diagnose model performance by determining the mismatch between their sensor features and their ground truth representation.**

While self-reports remain a mainstay for measuring mental wellbeing constructs like anxiety and stress, many studies in mHealth have posited alternative measures. Hovsepian *et al.*, proposed a new measure of stress in the wild, which involves a wearable device consisting of multiple biomedical sensors [359]. They found this measure to be a strong estimator of self-reported stress in the moment. Sometimes, physiological changes might not be captured in self-reports[330], but it is still valuable to characterize stressful episodes [360]. Prior work has provided evidence for these signals to trigger effective wellbeing interventions in field studies (e.g., heart-rate [361] and breathing [362]). Even though mental wellbeing constructs remain fairly subjective with respect to how they are experienced, perceived and eventually recorded [363], every kind of measurement is sensitive to different factors. For example, objective markers of physiological changes can vary with motion artifacts [359] and self-reports of psychological changes often obscure low-level details of the stressful episode [363]. **The presence of the semantic gap revealed in this work is meant to urge researchers to assess the imperceptible aspects of their ground truth measure while trying computational approaches to predict such constructs**.

7.5.2 Parsimonious Sensing

> For practical field deployments, the changing socio-technical landscape affects resource availability and privacy perceptions, which can limit researchers from conducting brute-force passive sensor studies with multiple complementary streams. Therefore, researchers should determine the smallest set of streams

that are semantically the most representative of the ground truth measure. Less is more if studies select sensors that provide features that reduce the semantic gap in predictions.

As new sensing platforms become commercialized and other interfaces like social media become abundant, researchers have a plethora of means to digitally infer their mental wellbeing. One approach to mitigate the semantic gap is to capture more ecological information that can help explain the high-level processes that influence the ground truth. What is evident from this study is that a single sensor stream is typically not robust enough to represent the different types of variability in ground truth. While offline sensors are skewed to represent behavioral changes (M_{pa}), online logs of virtual presence are better suited to represent social effects (M_{sm}). Therefore, a natural argument to reduce the gap between input features and target construct would be to deploy more sensors and track logs from multiple sources. In fact, combining multimodal features together can elicit new context-specific features [364, 275]. However, multimodal studies are challenging to deploy in the wild [111, 94, 113, 365], as they are expensive in terms of both instrumentation and recruitment. Moreover, additional sensors to capture the "reality" of a participant can introduce privacy concerns and generally overwhelm their experience [366, 367]. Instead, my findings suggest an alternative position to pursue parsimonious sensor deployments, or to make the most of limited resources to appropriately sense mental wellbeing constructs. I am inspired by Plötz's fifth postulate, "data rule, models serve" [358]. For instance, if deployments intend to measure ground truth through self-reports and researchers do not have access explicit sources of social signals (such as online activity or conversations), researchers should try to accommodate for social effects in offline sensors (as demonstrated by the Bluetooth beacons used in M^*_{pa} in subsubsection 7.4.1). Alternatively, if the study plans to estimate wellbeing with physiological changes then resources should be allocated to sense behavioral markers, such as movement and sleep. The existence of a semantic gap supports the idea of minimal sensing to predict wellbeing in comparison to conventional ideas of massive sensing. **Thus, my study demonstrates realistic approaches to adhere to paradigms like "small data" in (critical) data science [367, 368] and passive sensing [369], and the Occam's razor metaphor for parsimony in machine learning [370]**

In the meanwhile, more sophisticated methods to identify markers for mental wellbeing from passively sensed computational data have emerged [371]. Arguably, better feature crafting can help reduce this gap even with the same set of sensors. In this regard, the semantic gap serves two functions. First, it provides a guiding rail to engineer features based on domain-driven aspects of mental wellbeing ground truth. Second, it provides interpretability to models by encouraging researchers to inquire if their features capture psychological or physiological aspects of wellbeing. Moreover, the presence of a semantic gap calls into question the objectivity of machine learning/data mining to generate inferences. Since unobtrusive sensing can capture vast amounts of information, engineering this data can often yield spurious connections with the target variable [367]. My findings encourage more critical investigations of computational models to arrive at theoretically meaningful interpretations. Researchers need to resist the allure of viewing more passive data as a *Maslow's golden hammer* [372] — a tool to solve any problem. Over-engineering the "hammer" can result in finding spurious associations in the data [367, 373]. For example, does sensing physical behaviors actually predict stress holistically or does it merely describe its physiological aspects? Conversely, does tracing online content explain what an individual experiences or does it only reflect how they project themselves? Similar to other works that critique, yet advocate, employing machine learning for health and wellbeing [338, 374], **this study encourages researchers employing passive sensing to build models with deeper consideration of the domain and select sensors accordingly to avoid misrepresenting seemingly objective results**.

CHAPTER 8
SOCIOTECHNICAL CHALLENGES OF DEPLOYING PASSIVE SENSING FRAMEWORKS IN AN INFORMATION WORKER'S ECOSYSTEM

The previous chapters have discussed a variety of passive sensing frameworks that leverage different everyday digital technologies to clarify individual behaviors, social dynamics and organizational norms. Through these investigations, I aimed to enable a future of work that is more flexible and accommodating to different information workers. I repurposed digital technologies embedded in information worker routines to allow these frameworks to be easily deployed. My research has even discussed how to conceive minimalist, yet meaningful, passive sensing frameworks in resource constrained environments. The frameworks discussed so far require little manual input from workers themselves, promote inexpensive collective adoption, and do not disrupt the existing workflow. From this perspective, it appears that such frameworks can provide valuable insights at practically no cost. Given workplace power dynamics, however, this might be an illusion. **Hence, it leads me to a critical question, who are these passive sensing frameworks designed for?** Are these frameworks empowering information workers with a living record of themselves to break free from static measurements and improve? Or, are these frameworks arming employers to treat information workers as interchangeable commodities?

RQ V: *What are the sociotechnical challenges of deploying passive sensing frameworks in an information worker's ecosystem?*

Previous chapters demonstrated that passive sensing presents several opportunities to clarify worker effectiveness. Understandably, Passive Sensing–enabled AI (PSAI) has promising potential given that it is automatic, continuous, and unobtrusive. Although information workers are familiar with digital tools that count and track instances of their work[1],

[1]Common examples would be project management tools such as *Github* [375] and *JIRA* [376] which

PSAI distinguishes itself by collecting different kinds of data — peripheral and orthogonal to specific tasks — and algorithmically interprets these data to generate inferences of worker experience. I refer to these outputs as **experiential insights**. Therefore, I have focused on approaches that go beyond statistical measurement and incorporate increasingly complex machine learning to estimate individual's behavioral effectiveness [377]. These technologies promise to provide objective and precise insights into both performance and mental wellbeing [378, 111]. Admittedly, there are some positive outcomes of adopting PSAI. In concept, it stands to remove implicit biases in the workplace and make explicit many overlooked factors [379, 111]. That said, scholars and labor advocates also note problematic uses of PSAI that could harm a worker, as workers exist in a power asymmetry, where they may be disenfranchised [380].

Recent research has emphasized anxieties among data subjects whose data are used for making algorithmic inferences for work purposes (e.g., modeling past experiences to predict success) [381]. These concerns range from potentially exacerbated discrimination and compromised privacy expectations [382]. Moreover, tensions between supervision and surveillance in the workplace have been well documented [383, 384, 385]. Aloisi and Gramano rightly noted in their work that "Artificial Intelligence is watching you at work," given the emergent new practices of individual-level profiling, organizing, and monitoring, made possible by AI [386]. Unfortunately, some organizations are only letting IWs work remotely if they use passive monitoring, forcing them to relinquish their privacy [387, 379]. Adding to this concern, many commercially available instances of passive sensing for work are not designed for self–reflection or self–management and thus expect IWs to become data subjects of an obscure information flow has unclear benefits for their own growth, instead presumably catering only to the employer's interests [388, 389].

Albeit for the greater good of improving worker effectiveness, the findings from my research and related literature may encourage more oversight — a grave privacy concern that

expose worker activity to themselves and their peers.

underlines such social and ubiquitous frameworks [390]. Addressing some of these challenges extend beyond the scope of a conventional computer science dissertation. Complete solutions would require legal reform, driven by an inquiry into the civil liberties of workers. Yet, as a computer scientist, we have an opportunity to contribute to this discourse. I aim to inspect passive sensing frameworks to inspire more humane realization of these approaches. I believe, to prepare information workers for passive sensing in the future of work, we need to take a worker-centered view of applications and steer the development of passive sensing applications to reduce the imbalance within workplace power structures.

The studies in the previous chapters primarily assumed the role of the worker was only as a data subject. They provided their behavioral data to some passive sensing framework for further modeling. As is often the case in such research, data subjects were isolated from the model development and the ultimate application of their data. Instead, in these studies, I was particularly interested in how workers evaluate these frameworks as technologies they will adopt in the future of work. Understandably these information flows have other stakeholders but I choose to focus on the workers themselves as the data–subjects' voices are often missing from discourses around PSAI [391].In my interactions with workers, I referred to these technologies as PSAI [2] — short for *Passive Sensing–enabled AI* — to emphasize their predictive nature. **Study 1:** I first conducted an exploratory study with 28 information workers to learn the contexts in which they find PSAI appropriate and how they envision information flows that protect their best interest. This evaluation helped me hypothesize potential factors in a passive sensing framework that indicate a worker's willingness to accept the technology. **Study 2:** I then conducted an experimental evaluation of 1059 vignettes with 110 information workers to disentangle the best designs for passive sensing frameworks. In summary, the first study explains how passive sensing frameworks, as a suite of technologies, can be situated in information work for both enable and hinder worker success. The second study unpacks which passive sensing technologies are likely

[2]Pronounced "Psy" as in *psyche*

to find greater acceptance among workers.

Reflexive Considerations. I describe my positionality as a way to situate the values that shaped this research. My collaborators and I have conducted research in the past combining machine learning with passively collected data for digital phenotyping to support mental wellbeing. However, we have no stake, financial, personal, professional, or other, in any of the technologies used to inspire scenarios (Table 8.2). Yet, my research advances technologies like PSAI through novel methodologies as well as human-centered evaluations. In light of this, I consider myself an "insider" because this perspective critiques technology motivated by my own research and reshapes my own sociotechnical reality as a worker. My identity and experiences as a researcher also helps me construct meaning from my data and conceptualize my findings [392]. Broadly, this study is influenced by my interactions with privacy researchers, organizational psychology researchers, IWs, and other data subjects in digital phenotyping. I borrow Chancellor *et al.*'s term to describe myself as a "critical insider" [393]. I am at a unique position to bridge disparate views and approaches on the future of work by pursuing a worker–centered approach.

8.1 Study 1: Identifying the Contextual Norms of Using Passive Sensing to infer Mental Wellbeing and Performance

Information work, is notorious for having nebulous indicators of effectiveness, and by corollary, success [2]. This ambiguity presents an opportunity for PSAI systems that can not only model work activities (e.g., application time use, mobile distractions, work synchronization), but also model non–work correlates (e.g., sleep and movement) [80, 81, 83, 86, 95, 99]. Simplistically, the information flow of PSAI begins with behavioral data captured from the subject, which is then modeled by AI to produce inferences. In the reality of information work, such information flows are likely to be complicated by factors like how the data is collected, whom the inferences are shared with, and for what purpose.

Nissenbaum' *Contextual Integrity* framework states that to protect the interest of the data subject, new information flows must follow new informational norms [394]. This study aims to explain those norms for using PSAI for information work through two questions:

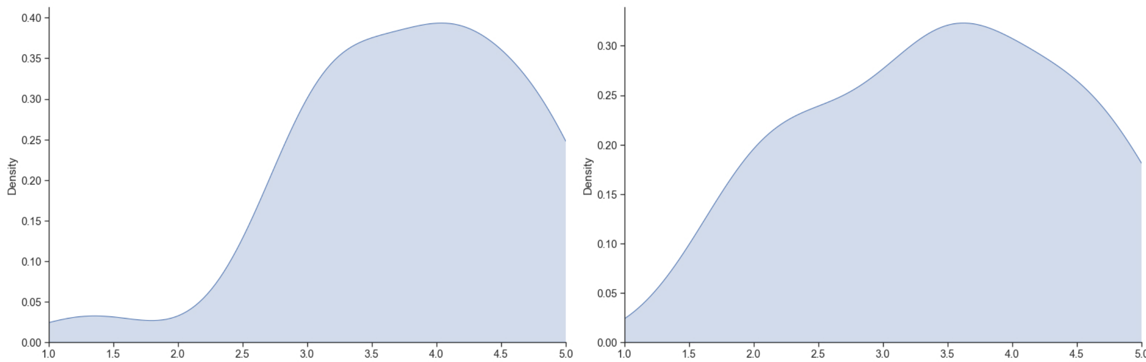
Norms of Appropriateness: What is the suitability of PSAI within IWs' expectations of algorithmic inferences of performance & wellbeing?

Norms of Distribution: When is it reasonable to share PSAI's inferences of an IW with other stakeholders?

I conducted scenario–based interviews with 28 IWs to highlight their perspectives on using PSAI to algorithmically infer their performance and wellbeing. I found that IWs envisioned powerful uses of PSAI but were aware of privacy intrusions and misappropriations. On the surface, this might appear as another paradox but the contrasting perspectives of supervision and surveillance can inform each other [395]. This study extends recent literature in HCI and CSCW that has critiqued algorithmic Human Resource Management (HRM) [381, 396]. Accordingly, we describe the norms for PSAI as guidelines for better information flows and improved regulation.

8.1.1 Methods

In this study I take a worker-centered approach to inquiring PSAI. In an institutional setting of information work, a worker is only one of the many different stakeholders. Studies on *Human–Data Interaction* describe data as common objects for all stakeholders to interact around [397, 398]. However, especially in this case, the data is not created by all stakeholders equally, nor are its implications uniform. When technology is designed without the benefits of the data subject we risk worsening the power asymmetry [399]. One method to tackle this growing asymmetry is by designing for the data subject as a primary beneficiary of a system that leverages their data [400]. Thus, I focus on the IW's perspective and investigate how they envision adopting PSAI in the future if at all.



(a) Attitude towards public surveillance (b) Attitude towards personal tracking

Figure 8.1: Distribution of participant attitudes. Higher values represent more acceptance. (a) *Public surveillance:* Opinion on expansion of surveillance to reduce crime and offences with a 3–item scale, (b)*Personal tracking: Experience with wearables, location tracking, social networks, etc. using a* $5--$*item scale.*

Participants & Recruitment

I recruited 28 IWs working in the U.S. and interviewed them between April and May 2022. I used both online and digital advertisements to recruit participants. I screened to ensure interested individuals had "work experience that involved cognitively demanding tasks to meet information-oriented goals, e.g.,, programming, marketing, engineering, accounting, management, etc.". Participants were required to have at least 2 years of work experience. I also required participants to have some experience working on–site so that they could consider PSAI in light of both the traditional and emerging work context. The participants represented a variety of roles including engineers, developers, analysts, and accountants. Participants prominently described their occupational sector as Information technology (IT), but the sample also reflected views from areas such as finance, consulting, manufacturing, healthcare, and libraries. 12 participants identified as female, 15 as male, and 1 preferred not to say. 17 participants were younger than 30 years old at the time of interviews. Participants completed a survey to report their attitudes towards public surveillance and personal tracking (adapted from [401]). Figure 8.1 shows that the participants leaned toward expanding public tracking (to reduce crime) and had a diverse set of experiences with technologies that track them in their personal life. Each participant was compensated with a giftcard worth $20 at the end of the interview. Table 8.1 provides a lookup summary of each participant along with their study identifier. Note, I did not explicitly analyze participants by the categories in Table 8.1. Inspired by similar studies [366], I have included these for epistemological accountability and to express the scope of my study.

Interview Protocol

Recruited participants consented to participate in one-on-one semi–structured interviews. All interviews were conducted by the first author and included one other author as an observer. Interviews started with open-ended questions to understand the approaches participants' organizations were using to evaluate their performance and wellbeing. Then, I provided a definition of PSAI rooted in personal tracking and an overview of its potential in the work context [409]. This was followed by a scenario–based comparison exercise to elicit rich perspectives on PSAI for workers.

Arguably, situating potential data subjects in actual behavioral contexts can help anticipate real behaviors. However, implementing a multitude of PSAI systems and conducting field studies is impractical. By contrast, leveraging scenarios that describe emergent use–cases can anticipate actual behaviors in new socio–technical settings [410]. This technique has been used in passive sensing to rapidly evaluate new application designs [411, 412] and understand privacy perspectives [413, 414, 415]. This approach has also made its way to studies on human–AI interaction [416, 417]. Park *et al.* have used this method to understand perspectives on general applications of algorithmic HRM [381]. Given my aim is to highlight norms, scenarios can be a powerful approach as "presenting users with scenarios that push social boundaries helps to uncover where these boundaries actually lie" [411]. As shown in Table 8.2, the PSAI scenarios I presented in my study were adapted from real systems for HRM. Each scenario outlined the information flow of the PSAI system — (i) how data is sensed, (ii) what inferences AI produces from the data, (iii) how the

Table 8.1: Participants summary by gender, age, race, as well as their role and occupational sector. *DND*: “Did Not Disclose”, *AA*: “African American

ID	Gender	Age	Race	Role	Sector
P1	Male	21-29	Asian	Research Engineer	IT
P2	Male	30-39	Asian	Developer	Finance/IT
P3	Male	21-29	Asian	Analyst	Finance/IT
P4	Male	30-39	Asian	Data Engineer	IT
P5	Female	21-29	White	Product Manager	Insurance/IT
P6	Male	21-29	Asian	Data Analyst	Insurance
P7	Female	21-29	Asian	Consultant	Consulting
P8	Male	30-39	Asian	UX Developer	IT
P9	Male	21-29	Asian	Research Assistant	Manufacturing
P10	Male	21-29	Asian	Accountant	Venture Capital
P11	Female	21-29	White	Scientist	Government
P12	Female	30-39	White	Developer	IT
P13	Male	21-29	Asian	Account Management	Retail
P14	Female	21-29	White	Technical Service	Library
P15	Female	21-29	Black or AA	Project Manager	Research
P16	DND	30-39	DND	Team Manager	Healthcare
P17	Male	21-29	Asian	Product Manager	(Did not disclose)
P18	Female	21-29	White	Product Manager	Consulting
P19	Female	21-29	Black or AA	Recruiter	Education
P20	Female	30-39	White	Researcher	Health
P21	Male	21-29	Black or AA	Engineer	IT
P22	Male	30-39	White	Software Developer	IT
P23	Male	21-29	White	Financial planner	Consumer Goods
P24	Female	30-39	White	Customer Service	IT
P25	Male	21-29	White	Business Analyst	IT
P26	Female	30-39	White	Consultant	Marketing
P27	Female	40-49	White	Director	IT/Sales
P28	Male	30-39	White	Portfolio Manager	Finance

Table 8.2: I designed PSAI scenarios based on contemporary technology. I refer to these in my findings via the labels here.

Label	Description	Adapted From	Reference
Sys 1	Uses CCTV cameras to observe different activities in a workspace. Analyzes physical activities to measure your performance. The HR will receive a report of your performance.	*CCTV*	[402]
Sys 2	Records the webcam feed of your PC. Analyzes your presence, expressions, and surroundings to measure your performance. Your manager will receive a report of your performance.	*RemoteDesk*	[403]
Sys 3	Captures screenshots of your PC activity at regular intervals. Analyzes PC activity to measure your performance. Your manager will receive a report of your performance.	*Interguard*	[404]
Sys 4	Uses custom sensor hardware to measure occupancy in different spaces at work. Analyzes the physical space use to measure performance. The HR will receive an aggregated report of workforce performance in different spaces.	*FM Systems and Freespace*	[405, 406]
Sys 5	Logs data from organizational communication (e.g., email, slack, or, calendar) and infrastructural systems (e.g., WiFi, Bluetooth, or access cards), Analyzes digital and physical activities to measure the organization's performance and wellbeing. The HR will receive an aggregated report of workforce performance and wellbeing.	*Humanyze*	[389]
Sys 6	Logs the time you spend on PC applications and web sites. Analyzes digital activity to measure your performance. Your manager will receive a report of your performance.	*ActivTrak*	[388]
Sys 7	Logs the time you spend on work applications (editing, communicating and scheduling). Analyzes work-related PC activities to measure your performance and wellbeing. You will receive a report of your performance and wellbeing.	*Viva Insights and My Analytics*	[407, 408]

inferences can be distributed. To improve elicitation, I showed participants two randomly selected pairs of scenarios. This approach was inspired by psychology literature that shows comparisons can be effective ways of rationalizing underlying features of an artifact by associations and contrasts [418]. The comparison of scenarios was not aimed to rate PSAI for HRM but only to initiate reflection. I also showed a third pair as a combination of already shown scenarios to provide additional rigor and clarity of preferences. To elicit perspectives, for every pair, I asked participants which scenarios they would resist to consent and which scenarios they would find useful. Note, the scenarios were only starting points and participants were free to reimagine PSAI as they described their preferences. For example, in certain sessions, participants only liked some aspects of a system but had problems with others. They had the flexibility of rethinking the scenarios. The interviews would continue with new emergent scenarios with the original scenarios only as reference. As such, the aim of the scenarios was not to show participants an exhaustive set of systems but rather provide a probe to help them appreciate the range of possibilities.

Interviews were conducted over Zoom and each session was recorded for transcription. Each transcription was scrubbed to remove any mention of the participants' employer or other identities of coworkers. Participants were informed that turning on the camera was optional. Interviews lasted between 40 minutes to 1 hour. This study was approved by the Institutional Review Board (IRB).

Data Analysis

I compiled all the transcripts and performed thematic analysis to synthesize patterns from the participants' perspectives [419]. Every transcript was carefully read and open–coded by at least two researchers. I was coded all transcripts. Throughout this process, we iterated the codes by meeting regularly to reconcile existing codes and identify new ones. After the codebook was completed, we performed affinity mapping to interpret and organize the initial codes into higher–level themes. This resulted in a three-level thematic structure. At the highest abstraction, the themes summarized IW perceptions of PSAI in terms of its effectiveness, concerns, applications for personal utility, and applications for shared utility. Given my aim to describe the norms of passive sensing, I reoriented and refined the themes as per the *Contextual Integrity* framework [394].

Contextual Integrity of Sensing at the Workplace. One of the classical approaches to evaluate privacy for passive sensing is to evaluate it by proportionality to existing activities [420]. In information work, project management tools such as JIRA are already used to disclose an IW's work activity to others on the project [376]. An IW might want to disclose their state of wellbeing to their manager to negotiate work breaks. In theory, this can be a compelling crutch to justify PSAI at work. Yet, it remains an open question if the algorithmic phenotyping of PSAI, introduces uncertain imaginaries that cannot be reconciled by existing work practices. According to Nissenbaum's framework of *Contextual Integrity*, user preferences for tracking systems are limited when privacy is considered intrinsic to the actors, spaces, or nature of information [394]. Instead, the adoption of systems must be studied by understanding the role of that information within the context of the user.

The contextual integrity framework is becoming increasingly significant to evaluate reasonable implementations of sensing technologies. Nicholas *et al.*, have illuminated attitudes toward personal sensing in the health context [421]. Similarly, contextual integrity has been used to explain adoption of tracking systems for public health [422]. Closer to my scope, a recent study by Adler *et al.*, described the norms of information flow for quantifying the stress of physicians in response to burnout [396]. Interestingly, in their context, workers felt that sharing information with a supervisor could be more valuable, than self-reflection, as supervisors had actual power to make changes to assuage their stress. Studies indicate that workers are willing to adopt ambient technologies if they enhance their well-being [423] and location tracking when it improves work efficiency [401]. Can we transfer these expectations to the algorithmic inferences provided by PSAI? Information work provides a unique setting. *This motivated me to interpret the themes from an analytical lens* *that reconciles the expectations of emerging technologies in specific settings*. We know that *Contextual Integrity* is upheld when the following information norms are maintained; (i) Norms of Appropriateness and (ii) Norms of Flow/Distribution [394]. As a result, I synthesize and scrutinize my findings based on these norms.

PSAI for work is different than PSAI for life Self-tracking can has a variety of benefits for a user like meeting physical health goals, managing finances, and even avoiding distracted driving for insurance [424]. In the aforementioned cases, this exchange of data is largely for personal benefit. But the same is not true for workers [425]. Before elaborating on the findings, I wanted to establish why information work presents a unique context for PSAI. IW attitudes were underpinned by existing work dynamics and expectations which made the adoption of PSAI systems at work distinct from those in personal life. When participants reflected on their use of personal tracking technologies (for fitness, sleep, and screen use) they were motivated by *"benchmarking"* (P28), *"hitting goals"* (P6), and *"tracking progress"* (P24, P28). Overall, these motivations aligned with visions of PSAI at work, to provide insights for self–efficacy and care. A key concern of using PSAI for personal tracking was data being used for advertising but this was perceived as a necessary transaction (*"I try to function in reality"* - P28). However, information work presents a unique context for using PSAI, with its distinct considerations. P14 articulated the overarching tensions that complicate the adoption of PSAI at work, *"On the personalized Fitbit, I am paying them to give me the insights. My request for that information outweighs my sensitivity for it [versus] personalized insights on technology driven by a company that is paying me to do work."*. Therefore, adoption of PSAI can be disincentivized by anticipated information flows and the existing power structures. Through the next section I elaborate on how IWs imagine the role of PSAI within the power dynamics of work.

8.1.2 Findings: *Norms of Appropriateness*: The suitability of PSAI within IWs' expectations of algorithmic inferences of performance & wellbeing

According to Sappington, the gap between actual worker behaviors and organizational perspective of workers describes an inevitably incomplete social contract that gives workers discretion but also limits organizational feedback [426]. This incompleteness can largely explain the motivation of PSAI at work [111, 137, 14]. The participants echoed the opportunities for PSAI–like interventions for their benefit, but they were also wary of the implementation of data collection and implications of inferences. Existing assessments of performance ignored the IWs' process, focusing just on outcome based *"statistics"* (P14) that provided a *"limited data view"* (P14). Instead, the participants had a more nuanced perception of their performance that could be reflected in work phenomena — such as break taking (P11), task-switching (P3), and availability demands (P19) — and non-work phenomena — such as their expressions (P24), sleep (P3, P7, P9), and physical activities (P13). Although wellbeing evaluations were not common, organization did provide resources (e.g., seminars, subscriptions to apps). The main complaint against these was the lack of individualized actionable information which made IWs feel their mental wellbeing was not actually valued nor was it important to the organization (P24). P3 exclaimed the missing link to be *"actual rubber to the road metrics, reaction and solution"*. Understandably, PSAI has promising potential given that it is automatic, continuous, and unobtrusive. However, efficacy in developing personal mindfulness does not sufficiently explain appropriateness in the information work context. This section describes how IW perceptions of appropriate PSAI were embedded in their attitude towards information work.

Effect on Job Consequence

PSAI systems provide indicators, which might be considered orthogonal to work specific tasks (e.g., your performance was moderate or stress was high). On one hand, participants found value in leveraging these insights to contextualize their experience with evidence and champion change. On the other hand, participants were anxious that these insights can be misappropriated to their own detriment.

> *"You've had these goals, you've had hit these hurdles. If you put that report in the context of this performance evaluation, I think together they're going to really have a significant impact on on your own professional and personal development."* - P6

The insights generated from PSAI can be empowering to IWs as it helps contextualize their work experience. P25 imagined such systems to support IW needs, *"I think with the data, it would help at least help you sit down at the table, so to speak."* Traditionally, workplace evaluations favor ends over means. As P2 puts it, *"the work which is getting done is what is counted, but how we achieve it is never logged in anywhere."* P5, like P6 (quoted above), believed that PSAI insights can complement existing evaluations after seeing *Sys 5*. Reflecting on one of her past evaluations, P15 claimed that PSAI could have been useful as a reference (*"let me give some numbers"*). It can give IWs a deeper understanding of their work patterns, present opportunities for learning effective work practices, and enable them to negotiate changes. P22 envisioned using *Sys 7* to request time off for their wellbeing, *"He can look into this report and it would be some kind of objective"*. Alternatively, P17 believed this data would be more persuasive to reorganize his work expectations, *"Having data that would support, I need a virtual assistant or we need to hire another PM or it's not feasible for me to run this many projects and run this team at the same time"*. P14 thought PSAI could support her in highlighting her role to others higher up in the organization. She said, *"This would end up benefiting us more because it would help others see how much we actually do and change the current stigma"*. Even at an aggregate level, it can help employers reflect on their organizational health. For instance, P25 found this a suitable approach for *"the company to be aware of work–life balance"*. Therefore, IWs find value in such systems when they can incorporate its insights into demonstrable change such as professional development or negotiation for wellbeing.

> *"Realistically, there is that concern that they're going to look at this big promotion and they're going to say, 'I don't know if he's going to cut it' "* - P3

The existing power asymmetry of information work environments always engenders concerns of privacy and subsequent misappropriation of their passively sensed data. Antithetical to the empowering aspects PSAI, P3 was concerned that his employer could tap into these insights to stifle their career progression, even going on to call one PSAI system *"destructive"*. P1 had a more straightforward concern, *"If my workday performance and how I work was released, it might affect how much I get paid."* Participants like P4 were uncertain about consequences of others stakeholders using this data but this uncertainty made them anxious. These concerns stemmed from the perceived lack of control over one's data in the organizational context. "I download the app, the information is captured and then it goes to someone else. That's the objection." (P27). Furthermore, implementing PSAI like *Sys 1* and *Sys 4* — which are embedded in the physical infrastructure — can create an austere situation, where IWs might feel that their choice to consent could affect their employment (P5, P11). In fact, some participants felt that the very decision to use such systems for deeper surveillance can reflect an organizations' own values (P12). Eventually, such uncontrolled and imbalanced deployment of PSAI can detract IWs from choosing to work in such companies. However, even that choice is a function of job precarity in that sector. As a result, development of PSAI systems needs to be aware of the socio-economic conditions of employment.

Respecting Work–Life Boundary Management

Workers' preferences for work–life boundary management reflected their perceived control of privacy in PSAI systems, but also highlighted their expected value from the system. Post–COVID-19 pandemic, new emerging work practices are allowing many IWs to work remotely either in their entirety or on certain days of the week. Moreover, given the ubiquity of personal laptops and mobile phones, it is commonplace to bring some work home.

Although, non-work can influence work experiences, some workers found sensing beyond work invasive and irrelevant to improving work. Yet, some workers also believed that sensing non-work could be less consequential to their jobs and more holistic for reflection.

> *"If I'm going to the office, I will probably agree to do that. But if I work from home [...] I don't want that to record anything in my home that's maybe not work-related. "* - P8

Different workers have different approaches to their work–life. Some demarcate the segmentation between the two using physical aspects. A common understanding is segmenting work–life based on the space the IW finds themselves in. In the quote above, P8 was willing to consent to PSAI if it is contained to their workspace. With more interleaving work–life practices, space is not the only indicator of work–life separation. Organizations often provide workers with work–specific devices or enforce a logical separation between work & personal profiles. For example, P5 thought *Sys 5*, which logs applications and browsing, was reasonable because she did not do personal activities on her work machine anyway. Although this might be to ensure security of organizational data, it also provides another method for IWs to segment work–life. P17 noted that he was willing to allow PSAI systems on his work device, *"But if it's a personal device and I'm doing work on, absolutely not."* P1 said *Sys 2* was a *"violation of personal space"* because the webcam could capture their home environment. Understanding these constraints can help describe the limits within which privacy can be preserved. It is also worth noting the concern of some participants who believed that preserving the work–life boundary for PSAI made it more useful (P13, P16, P21). Similarly, on different occasions, both P9 and P10 stated the focus on the work context was more *"accurate"*. *"If we can achieve only tracking the work applications that will definitely improve the efficiency and avoid a lot of other privacy arguments, if there's any there"*, said P8. Thus, for certain IWs, the work context is not only more private but can actually be more useful.

> *"Maybe on a Fitbit watch or something wearable rather than my computer itself, because*

I don't like people seeing what I'm doing on this computer" - P7

P7 presented an alternative viewpoint that shows work–only restriction of PSAI can elicit concerns about job consequences (subsubsection 8.1.2). In fact, depending on what kind of data is being sensed an IW might consider the privacy of their work activities to outweigh that of activities outside of it. Devices distinct from the work context can be considered more reasonable for sensing. P1 even described a greater willingness to accept a PSAI system provided by a third–party because of the apprehension that something from their organization can be misappropriated by HR. Again, the shifting of sensing to non-work devices and concepts is not only determined or shaped by privacy decisions, but P18 found other work–specific PSAI to be limited in *"the world of working from home"*. P6 felt that PSAI like *Sys 4* could be more valuable. He said, *"It would give me a true reflection of of how I work, it would give me a true performance evaluation report that I can actually make use of."* Thus, PSAI systems that model phenomena outside work could provide the opportunity for an IW to interrelate all aspects of their life and improve as a totality.

Preservation of Flexibility

Choosing where to work is not the only freedom IWs have in determining their work styles. IWs often enjoy a broader sense of flexibility where they are rewarded and evaluated for outcomes. Unlike other forms of labor, an IW is not as heavily scrutinized on time–tracking. "Brain work" is often hard to quantify and therefore workers can approach work tasks at their own rhythm. I found that the participants suspected this flexibility could be hindered with PSAI systems, even if their employment was unaffected or their work–life boundary was secure.

> *"I'm very flexible in how I work. I like to get things done on my own time. Sometimes that means I carry over on the weekend. And sometimes that means I just do work nine to five. I'd rather just keep that on my own and how we get things done rather than having some kind of tracking."* - P1

Among others, P1 felt that PSAI inferences might be reductive in quantifying varying work styles. In reference to IWs that work in bursts or "sprints", P9 said, *"it can it can adversely affect people who do not progress in a linear manner."* Similarly, P20 anticipated a simple case of regimenting where PSAI's inferences would force her to work specific hours instead of simply being judged on her output. For IWs like her who work from home, these systems could disrupt how they choose to interleave their work–home responsibilities. P18 noted that *Sys 1* could penalize her behaviors that do not look like work, but actually are, such as when they *"do laps around the office"* or *"lay on a beanbag chair."* P7 believed this to be the case when PSAI was limited to *"just PC stuff"*, such as *Sys 3* and *Sys 6*. More generally, P28 believed that PSAI confined to work applications reinforces an *"older view"* of work that thinks *"you've got to be in a place to be able to do a job."* With more IWs opting into remote or hybrid work options, these systems can be considered regressive. However, expanding sensing might not be the solution either. IWs like P1 found that tracking ecological factors like movement and space might not generalize to *"varying situations"* and could render false negatives. Meanwhile, P6 preferred work–specific tracking over ecological ones because, *"I could skew the data in favor of my performance being better than it actually is"* (referring to screenshots taken by *Sys 3*). As a result, it would free him to work as he likes. These findings indicate that the very presence of PSAI could establish expectations of a rigid work style and discourage pluralistic approaches to work.

> *"It could become dehumanizing. It could be become a little bit robotic, like in a way like I can only perform it seven percent today."* - P19

The threat to flexibility posed by PSAI can not only restrict activities but also lead to worker distress. In the quote above, P19 alluded to feeling further commoditized because algorithmic inferences tend to convert nuanced, complex human experiences into streams of numbers. P3 described enrolling into such a system as *"a little intimidating"* and feeling like a part of a *"cold, hard, big institution"*. These impressions were likely due to perspectives of PSAI as a tool for reducing the complexity of a worker's experience into performance metrics. P10 felt that *"continuously monitoring for what you do [...] could affect your job more"* and P20 even thought it could be *"distraction and counterproductive"*. Similarly, P5 described that *"I don't like being overanalyzed [. . .] I would be less likely to produce good work."*. Other participants like P6 and P11 also alluded to the fact that PSAI systems can exacerbate the Hawthorne effect caused by supervision [427]. It could even lead to negative consequences to an IW's affect. P24 was concerned that the continuous monitoring required by PSAI could be *"stress [her] more"* and P10 felt it would *"put a lot of pressure on [him]"*. These perspectives mostly arose from discussions on performance measures and not so much on wellbeing inferences. However, to keep up with a camera based system like *Sys 1*, P8 felt they would need to compromise their wellbeing by reducing breaks and socializing at work. Even when a worker might not lose sensitive data, the mere presence of these systems can impact their work effectiveness. This risk is ironic for systems that aim to improve worker wellbeing & performance.

8.1.3 Findings: *Norms of Distribution*: Reasonableness of sharing inferences from PSAI with other stakeholders

Information work is inherently collaborative in nature. Collective knowledge supports IWs in their day-to-day and during important career junctures, such as evaluations and promotions. The participants explained that coworkers were *"disconnected"* from others' challenges (P21) and the lack of awareness of each other's state led to disruption in work (P26). On the contrary, IWs needed to personally check on each other's wellbeing (P5, P14) and felt that hybrid work was diminishing their ability to maintain this practice (P25). Some PSAI systems could potentially smooth out organizational workflows by pooling of behavioral patterns [428, 429]. Arguably, such existing practices would present possibilities for IWs to share estimates from PSAI within their work network. In this section, I describe the different paradigms that motivate an IW to share and the conditions within which they think sharing should transpire to protect their interests.

Paradigms for Sharing

Sharing knowledge in an IW's workplace is essential for seamless communication and information flows. PSAI systems might develop insights on a worker passively but how and where that information is distributed needs to be a deliberate process. Here, I describe the network of stakeholders within which a particular IW might want to share the insights provided by PSAI.

> *"Sometimes people don't know how to manage the stress they have at work. I've seen that with a couple of people. They don't know to how to escalate that or to communicate that up. You know, and sometimes supervisor doesn't know because they've never been able to see it." – P25*

A common form of distribution described by the participants was the value in *one-one* sharing for personal improvement. Others echoed P25 in using PSAI to better explain their work context to their managers (e.g., *Sys 2*, *Sys 3*, and *Sys 6*). This perspective was often described analogous to existing workplace practices, i.e., *"I work with them"* (P15), *"they know more of your day-to-day"* (P7), and *"understand the way that thought processes work or deep thinking happens"* (P18). What is apparent to be important in this sharing flow is that the manager should be viewed as a stakeholder with real expertise or valid opinion on an IW's work behavior. Participants suggested alternative experts as well such as, senior collaborators (P7), or advisors and mentors (P6). Sharing PSAI inferences could also be seen as necessary because viewing them in isolation could cause harm. About *Sys 7*, P20 pointed out that she *"would have to use it with some sort of guide or some sort of other coach in order to use it in a kind way rather than in a punitive way."* On inspecting P20's sentiment further, it was clear that her preference was not for validating estimates but rather for *"someone to tell me that what is happening is normal in this moment or common."* Note, however, that these one–one flows still reflect the power asymmetry of information work as in most of these cases an IW is required to share information with those above them in the hierarchy. Refreshingly enough, I also found some perspectives that go against these expectations. Both P16 and P28 were willing to share their PSAI insights with those they mentor. Thus, a key component in this paradigm is the presence of stakeholders who have the know-how to reappraise the inferences produced by PSAI and in turn generate more holistic feedback for IWs.

> *"I know if I hear that our team is doing well, but I know I'm not up to standard and what I should be meeting in terms of the team... that actually pushes me a little bit more"* - P13

Another sharing metaphor that emerged from my findings was to share information for comparisons and coordination in a *many-to-many* fashion. The commonly stated purpose was to share PSAI insights to regulate their activities in accordance with the coworkers they aspire to resemble. Akin to P13, P22 felt that seeing others could help them aspire for better work processes. Of course, these many–to–many transactions need to be mutual as P28 said *I'd be willing to share my price in order to get that back."* P10 also wanted to be compared but clarified that it did not have to *"be a specific number"* suggesting abstract methods of comparison. P7 wanted to view aggregated insights for *"different types of roles like people that are managers or [...] how much are the analysts doing?"*. Thus, this kind of aggregated benchmarking against other peers can also help an IW identify "normalized" patterns and improve social awareness within organizations. This could be cathartic but could also nudge an IW to understand that certain challenges might be a function of a poor workplace, that are less "mutable". Aside from self–regulation, some participants believed this form of sharing could help redistribute workload. P7 envisioned sharing PSAI estimates of her stress with coworkers so that, *"they can at least take over some of the easier work."* To complement this P25 wanted to know how the people he was working with were feeling and support them during junctures of low wellbeing (*"it beats them up"*). Similarly, P18 thought this kind of sharing could help manage work in her team when one of her coworkers was *"having a hard time personally"*. An extension of the *many–many* paradigm discussed earlier is to share data completely anonymously for gross aggregation (e.g., *Sys 4* and *Sys 5*). While P4 claimed the lack of specificity in such a paradigm would make their privacy more protected, P20 and P26 believed it could actually serve an altruistic purpose. She said, *"I would certainly consent to that if my individual data were to be consolidated with others because I think that there would be a purpose to that."* Although this might not directly benefit the IW who is a data subject, this could lead to eventual collective benefit, such as cultural change in the concerned organization. Therefore, sharing within a finite network of coworkers can help an IW make more sense of the inferences they receive and leverage the support of coworkers.

Conditions for Sharing

In information work, some information about an IW is constantly available for all coworkers to see. It can be momentary information such as availability or work–specific information such as which documents they worked on. In regards to this kind of information, an IW's state can be visible to others unconditionally. However, with PSAI insights the participants preferred more intentional sharing. Particularly, IWs referred to their ability to negotiate, perception of stakeholder roles, and accountability of the system. This section expands on the factors that can inform flows of sharing.

> *"It feels a little vulnerable to just then send the metrics off to somebody without having a chance to add my own interpretation, just leaving it up to their interpretation"* - P15

In the quote above, P15 wanted to assimilate and communicate her understanding of the algorithmic insights before passing it over to another stakeholder. Similarly, P11 said, *"I feel like I have got some time to adjust and then sharing would make me feel more comfortable"*. P7 remarked that she would share the insights if *"it needs to be escalated"*, a common way of describing work issues in information work that need more attention. *"I would feel better, I would feel more in control,"* said P14, when reflecting on the possibility of personally appraising the PSAI insights first. Without this agency, an IW might feel over–scrutinized. P20 even exclaimed that this kind of sharing can be seen as *"manipulative"* but that social discretion allows her to fulfill any necessary disclosures to her supervisor. Essentially, IWs need some room to negotiate the algorithmic inferences of PSAI, before they can be distributed any further. Thus, the control that IWs seek is not only limited to whom the data is shared with but how and when it is shared.

> *"This is something I can learn, adapt and improve myself and also talk to my manager so we can work together to get better. But if HR sees something then I am not sure how he or she will respond."* - P8

The ability to negotiate the PSAI insights does not necessitate that newly generated information can be shared with anyone. Participants had varying attitudes towards different stakeholders in the information flow because of anxieties that the insights could be used against them. Several participants shared P8's view that their manager is preferred over the HR as a receiver in the information flow (P7, P15, P19, P22). Ironically, I found strong resistance to HR being involved in systems designed for HRM. The responses to these situations were plain and clear, *"I don't want HR to be measuring anything"* (P17). HR as an entity seemed to foster a negative connotation, described as a *"bad word"* (17), *"scary"* (P23), or *"ominous"* (P15). Both P13 and P24 anticipated they would be worried about what HR could interpret from PSAI or question them for. Aside from the social connotation, P22 thought HR was too *"far removed"* from the work context and P23 thought this was not a part of HR's role. Alternatively, not only was a manager more relevant to an IW's functioning, certain participants also noted that it was actually their manager's primary function to improve their work experience. P13 even went on to say that they would rather have PSAI directly send insights to managers, *"I don't really care about how well I'm doing in terms of performance, I feel like that's more of a important measure for my manager."* P4 and P10 described that managers could use these measures to coordinate work. By contrast, other participants expressed greater concern because of managers potentially micromanaging. *"I also like the level of separation from my manager..,"* said P5, who would prefer the insights being shared with Human Resources (HR). Similarly, P9 mentioned the trust deficit between him and his manager made it challenging to foresee favorable outcomes of sharing PSAI insights. Interestingly enough, P4 thought that HR should be the ones that educate managers on good practices for using PSAI. Therefore, the perception of the functional roles of different stakeholders can determine an IWs willingness to distribute their PSAI insights.

> *"Even though System X appears to capture more of what I'm actually doing. I'm personally willing to fork over that if I understand the details of it."* - P28

In the backdrop of the above anxieties, IWs like P28 expressed a need to understand how the entire PSAI information flow was setup. An improved understanding can help an IW anticipate the consequences of misappropriating these insights. P5 called out for greater disclosure and transparency, *"Some way to describe the limitations of the system would make me more comfortable with a system."*. In fact, some participants noted that they would be indifferent to PSAI they consider ineffective and therefore invaluable in the information flow (e.g., both P1 and P7 thought that space usage was a poor measure). Besides the mechanics of PSAI's collection and inference, P9 and P20 called for explicit disclosure of the stakeholders who could access these insights. P1 was skeptical about who manufactures the PSAI, *"If they were using a product promoted by Apple to do this, I'm going to be more OK because I know this data is not going to go back to HR."* P11 urged that the flow of information needs to established within organizational policy, *"I think there has to be an agreement of purpose or expectations."* For instance, participants with smaller teams were concerned that many–to–many sharing can lead to negative consequences for their job role or employment (P1, P2, P21, P22). Participants also valued awareness of who else was sharing or how many people were being aggregated. As P16 noted, *"given that the culture was such that everyone was having an openness to the material and they felt comfortable with it and it made sense to everybody"*. Therefore, IWs tend to expect clear notice and guidance on the scope of distribution in the PSAI system.

8.1.4 Summary

Passive sensing can be a powerful tool in enabling AI to infer worker performance and wellbeing. The use of such algorithmic evaluations for HRM in information work may not be widespread but is on the horizon. By investigating worker perspectives, my research discovers the norms that Passive Sensing–enabled AI needs to adhere to to maintain contextual integrity, while inferring effectiveness of Information Workers. I highlight factors specific to information work that can inspire appropriate information flows of evaluating IWs with PSAI and appropriate methods of share these information flows with others. This study thus helps to envision new worker-centric implementations of PSAI that do not breach their self-interest and dignity while also promoting their prosperity.

8.2 Study 2: Cost Benefit Analysis of Passive Sensing for Information Work

For any application or tool, people are likely to adopt it if the value it creates outweighs the cost of using it. For PSAI a central assumption, which is also its central promise, is its unobtrusiveness. PSAI can be continuously "on" without any effort from the user [13]. Further, with everyday digital technologies, we do not incur any additional overhead for purchase or installation of sensing infrastructure. These assumptions have led computer scientists to believe that passively collecting any data is reasonable, given that it does not cost anything. However, this assumption fails at the workplace due to two key challenges. First, the new perspectives on privacy assert that one's data itself is of value and therefore must be considered a cost [430]. Second, due to asymmetries at work, the information worker might be the source of the data, but may not receive the benefits [399].

> "The application fails because it requires that *some people do additional work*, while those people are *not the ones who perceive a direct benefit* from the use of the application." — Grudin [431]

This quote refers to traditional CSCW applications, such as groupware [431], but still applies to PSAI for the work. From the perspective of the information worker, the equivalence between data contributed and the value received is unclear. In fact, an employer can entirely dispossess a worker from their data [432]. In light of these flaws, its natural to ponder if PSAI systems are acceptable to their data subjects, the information workers. When studies on new PSAI innovations at work discuss ethical and societal implications, they mostly hinge on cautionary notes of informed consent. However, shifting the onus on the information worker often ignores the imbalance at the workplace. Not all information workers would have a say on this matter, especially given the precarity of work [425].

PSAI can be designed with a variety of information flows in mind. We already learned in the first study, that not all these implementations are equivalent. In this study, I investigated the which components of PSAI are information workers more willing to accept. The aim is to tease apart the idea that these technologies represent a monolithic representation and instead distill the valuable aspects separate from the counterproductive ones.

8.2.1 Study Design

It can be challenging to obtain comparative insight on worker preferences when technology is deployed in highly consequential situations, such as the workplace. Not only can it be expensive, but it can raise ethical concerns. To overcome these hurdles, I analyzed PSAI with the *Experimental Vignette Method* [433, 434]. A "vignette" is a scenario that reflects specific features or components of a technology, policy, situation, or interaction. Scenario based experiments have been used in the past to study algorithmic management at work [416]. The experimental angle of this method involves the presentation of a series of vignettes to participants for evaluation. These vignettes are carefully modified across certain components and in effect akin to a factorial survey that helps simulate real world conditions [435]. To study PSAI, I modified components that represented hypotheses I derived from literature and further refined from the previous study.

Hypotheses

To design my study, I initially drew from the Technology Acceptance Model (TAM) to predict the willingness of workers to be data subjects for PSAI. The TAM framework was first proposed to determine the acceptance of technologies for improving performance at work [436]. Additionally, TAM has been used in a variety of different settings including e-learning, cloud computing, virtual reality, and Internet-of-Things [437]. Traditional analysis with TAM involves two antecedents of acceptance, perceived utility and perceived ease of use. Arguably PSAI do not involve explicit use and lead to minimal interaction burden. Classical interpretations of TAM would consider PSAI acceptable. However, looking towards the theory of *privacy calculus* we can identify a different cost [430]. As per this, perceived ease of use in TAM can be replaced with perceived privacy risk [438]. Based on the findings from the previous interview study, I learned that some of the anticipated costs go beyond data privacy risks, and extend onto job consequences (section 8.1). Taken together, I believe a PSAI solution with more perceived utility and less perceived harm is likely to be more acceptable.

Type of Sensing. The first piece required to engineer a PSAI system is the sensing component. Different systems can leverage digital traces from different sources (Table 8.2). The CCTV cameras at work can be considered passive sources too. Over time, information workers have accepted such cameras as the norm. However, it can be closer to PSAI when that CCTV feed is input into a machine-learning models to provide metrics of work effectiveness. After remote work, the more concerning usage of visual feeds was actually the prospect of an employer tapping into the inbuilt camera on a worker's machine. *RemoteDesk* is an example of such a technology [403]. Charbonneau and Doberstein, have shown that people perceive the intrusiveness of camera–based applications differently from fitness trackers [439]. In comparison to other scenarios, when participants from Study 1 were exposed to *Sys1* or *Sys2*, they often described it with verbs, such as "seeing" and "watching" (P4, P7, P23, P28), instead of "tracking" or "recording". In terms of pure functionality even other technologies are digitizing and storing human behavior. The intrusive perception of these devices motivates the exploration of novel sensing sources. Research studies — including my own — discuss a variety of other sensors that can be used for passive sensing at work, such as smartphone screen use, bluetooth beacons, and even language online. The first study also made it clear that IWs felt that certain sensing sources could actually be more meaningful than others at indicating work experience. Some participants expressed that digital work applications could be an appropriate way to harness traces. This idea did have detractors though. Among others, P18 stated, *"the amount of time I'm on my computer or the amount of time I'm responding to emails to me isn't an accurate representation of productivity."* Alternatively, other streams such as physical activity and language might not appear as tightly coupled with work. Yet, workers considered these orthogonal correlates "true reflections" (P3). Therefore, it is reasonable to hypothesize that workers will favor the use of certain sensors in PSAI:

H1a. *Type of sensing stream is associated with perceived utility*

H1b. *Type of sensing stream is associated with perceived harm*

To test these hypotheses, I compared PSAI systems with 3 different sensing modes: (i) digital time use (e.g., time spent on an application, (ii) online language (e.g., sentiment and tone on communication platforms), and (iii) physical activity (e.g., number of restful breaks from a wearable).

Scope of Sensing. Historically information workers have been able to avail flexible work routines. They get the freedom to decide when to work and where to work. Naturally, the degree of freedom might vary across roles and organizations. This flexibility has blurred the lines between contexts when a worker is working (e.g., office) and other general situations where they could also be working (e.g., after-hours at home, at a cafe on the weekend). Simply deploying passive sensing frameworks without regard can heighten anxieties related to unchecked surveillance [432]. In subsubsection 8.1.2, I had described the various different opinions IWs had regarding limiting PSAI to work or allowing it a broader reach. While sensing beyond work could be a privacy risk (P17), sensing specific to the work context could be more sensitive to their career (P7). Workers also varied on the utility of different scopes. P25 felt that a broader scope could highlight their holistic needs, whereas P9 and P10 found work–specific scopes to be more relevant to their role as workers. The varying scope can not only change the traces that a PSAI system would capture, but also the way its output will be interpreted. I want to test if these distinctions explain the overall perceptions of PSAI systems.

H2a. *Scope of sensing is associated with perceived utility*

H2b. *Scope of sensing is associated with perceived harm*

To test these hypotheses, I compared PSAI systems with 2 different scopes: (i) work (e.g, work application tracking, work communication, or occupancy sensors) and (ii) general (e.g., personal application tracking, personal social media, or wearables).

Type of Insight. Any PSAI system will process the data to infer a target. Contemporary systems have typically centered on providing performance–based measures. Exceptions like *Viva Insights* present wellbeing insights alongside performance [407]. Abraham *et al.* found that workers are more likely to accept monitoring systems that promise an increase in work efficiency — producing more output in the same time [401]. In contrast, Cheney-Lippold found that technologies that monitor efficiency tend to burnout workers and constrain their day-level activities [440]. The need for performance insights arise from organizational incentives and social contract between an employee and employer. Recently, information work has started paying attention to mental wellbeing. We are witnessing a rising trend where organizations are conducting seminars, appointing specialized officers, and even offering mental wellness apps. However, mental wellbeing has not received the same individualized attention as performance. Performance evaluations in information work have been refined and embedded into the worker's life–cycle. *Key Performance Indica-*

tors and *Performance Reviews* are common place, but mental wellbeing is often addressed through nebulous actions. Participants in Study 1 echoed the feeling of not being cared for enough and a disconnect. These organizational interests are likely tied to preserving human capital. P7's perspective illustrates the different views on mental wellbeing, "Every month they send out the email, [...] OK, those are the best days because my mental well-being is usually the best on a days they send it out." Inferring a worker's health also presents new challenges. In the past, workers have expressed resistance to applications that track their physical health because of anxieties related to insurance possibly because of anxieties related to its effect on their insurance [401]. It remains unclear if workers would actually prefer insights on mental wellbeing. P28 was skeptical of the value of computationally estimating his mental wellbeing because "that kind of thing is so personal". Accordingly, it is yet to be learned if PSAI would be more valuable to workers if the kind of insight received by them is different from performance.

H3a. *Type of experiential insight is associated with perceived utility*

H3a. *Type of experiential insight is associated with perceived harm*

To test these hypotheses, I compared PSAI systems with 2 different types of insights: (i) performance, (ii) mental wellbeing (specifically *stress*).

Sharing of Insight. The asymmetry of work–based power structures inherently carry a critical risk. The insights of passive sensing can often be consumed by someone who is not the data subject [400]. Ideally, the insight of PSAI should affect an IW directly. In reality, it is possible that the insights generated by PSAI indirectly impact IWs through the organizational decisions of others, such as managers. The indirect flow is often designed to benefit organizational interests, e.g., to reorganize work within a team. Prior research indicates that people vary in their privacy concerns when comparing individual and collective benefits [441]. Information work relies on collaboration, communication, accountability, and dependency. In Section subsection 8.1.3 I had discussed various paradigms within which

IWs envisioned the use of PSAI. Participants had reported that they always want to be involved in the flow as a receiver, but could imagine forwarding insights to other coworkers. Some appreciated the role of the manager in enhancing their own work effectiveness (P15, P17, P18, P25). Others felt a *trusted–other* like a peer, coach, or mentor could help explain the insights (P6, P20). Additionally, these insights could also be contributed to a collective aggregate to keep workers updated (P13) and help the smoothen work-flows (P22). Together, all these potential uses motivate the last set of hypotheses:

H4a. *Sharing of experiential insight is associated with perceived utility*

H4a. *Sharing of experiential insight is associated with perceived harm*

To test these hypotheses, I compared PSAI systems with 4 different sharing paradigms: (i) self only (nobody else receives the insights), (ii) self + manager (iii) self + trusted other, and (iv) self + aggregate. Note, in variations (ii-iv), I specifically studied instances where the insight is shared 1-week after the worker has received it themselves.

Table 8.3: The vignette text was generated based on randomly selecting values for $H1$ and $H2$. Every participant was shown the baseline vignette where $H1$ =“Visual” and $H2$ =“General”.

H1	H2	Vignette Text	Explanation	
			Input	Process
Visual	General	The AI system can analyze video captured from your work computer’s webcam. It will extract facial expressions (e.g., attentive, stressed, enjoying) and will NOT store any identifiable images.	This PSAI system will use the primary camera of your PC. While the PSAI is running, the camera will remain ON. The camera can be an inbuilt camera or an external one.	This PSAI system will use Computer Vision and Machine Learning to learn your cognitive state based on subtle and complex facial actions. An example of what the PSAI will store ...[sample table with values for timestamp, “focus”, “distracted”, “calm”, “energetic”]. The system will NOT store any image or video of people or surroundings.

Table 8.2.1 Continued.

Online Language	Work	The AI system can analyze text retrieved from work-related communication platforms such as Slack, Teams, and Email. It will extract psycholinguistic attributes from your posts (e.g., frequency of positive emotions) and will NOT store any readable text.	This PSAI system will analyze the text you input into your work-related communication apps. The PSAI could include email (e.g., Outlook), instant-messaging (e.g., Slack), work-social media (e.g., Yammer).	This PSAI system will use Natural Language Processing to learn different figures of speech in the communication text (e.g., frequency of positive phrases). An example of what the PSAI will store: An example of what the PSAI will store ...[sample table with values for timestamp, "positive", "negative", "anger", "achievement"]. The system will NOT store any raw text or nouns.
	General	The AI system can analyze text retrieved from communication and social media platforms including Facebook, Twitter, and Search. It will extract psycholinguistic attributes from your posts (e.g., frequency of positive emotions) and will NOT store any readable text.	This PSAI system will analyze the text you input into any of your communication apps. The PSAI could include email (e.g., GMail), instant-messaging (e.g., Messenger), social media (e.g., Twitter).	[Same as above]
Digital Time Use	Work	The AI system can analyze your screen time on work applications such as those used for communication, development, design, documentation, and presentation). It will extract engagement measures for different categories and will NOT store any application names, titles, or content.	This PSAI system will analyze your interactions with work-related applications. The PSAI will be restricted to application categories such as work communication (e.g., Outlook, Teams), programming (e.g., VS Code, Github), and documentation (e.g., Word, Excel).	This PSAI system will use event logging to learn the time you spend on different categories of applications and how you use it (e.g., typing, clicking, scrolling). An example of what the PSAI will store ...[sample table with values for timestamp, "category", "app in focus", "mouse move", "keystroke count"]. The system will NOT store any content, such as file name, application name, or typed text.
	General	The AI system can analyze your screen time on computer applications including those used for work, entertainment, browsing, and gaming). It will extract engagement measures for different categories and will NOT store any application names, titles, or content.	This PSAI system will analyze your interactions with your applications. The PSAI will be cover a variety of applications categories including entertainment, programming, and browsing.	[Same as above]

Table 8.2.1 Continued.

Physical Activity	Work	The AI system can analyze movement patterns using occupancy and presence sensors embedded in the office infrastructure such as doors and desks. It will extract frequency, intensity, and rhythm of different physical activities and will NOT store any identifiable locations.	This PSAI system will use sensors embedded in the work space, such as access doors. The location of the sensor can indicate the activity you are involved in (e.g., at your desk or away). These sensors are used for security, maintenance, and analysis of space usage.	This PSAI system will use activity recognition to learn the time you spend on general activities (e.g., sitting, moving). An example of what the PSAI will store ...[sample table with values for timestamp, "activity type", "duration", "event count"]. The system will NOT store any location data.
	General	The AI system can analyze movement patterns using motion and physiological sensors in your smartwatch. It will extract frequency, intensity, and rhythm of different physical activities and will NOT store any identifiable locations.	This PSAI system will use sensors in your smartwatch. These sensors are used for providing different interactions (e.g., automatic screen on/off), but also to track your physical state (e.g., resting, or intense activity).	[Same as above]

Table 8.4: The vignette text was generated based on randomly selecting values for $H3$. Every participant was shown the baseline vignette where $H3 =$"Performance".

H3	Vignette Text	Explanation: Output
Performance	With this data it can estimates your job performance on a scale of 1-100 at the end of every day.	The PSAI will provide daily insights based on different aspects of performance. You will see a single score (1-100) that reflects (i) the quality with which you perform assigned tasks and (ii) the quality with which you perform additional unspecified tasks related to work.
Mental Wellbeing	With this data it can assess your stress on a scale of 1-100 at the end of every day.	The PSAI will provide daily insights based on different aspects of mental wellbeing. You will see a single score (1-100) that reflects (i) the stress you experience from external demands and (ii) the anxiety you experience thinking of future events.

Table 8.5: The vignette text was generated based on randomly selecting values for $H3$. Every participant was shown the baseline vignette where $H4 =$"Self+Manager".

H4	Vignette Text	Explanation: Users
Self	You will be able to view the system's assessment of you everyday and reflect on long term trends.	The insights from PSAI are only available to you. By viewing your insights, you can learn more about how you work. This new understanding can identify opportunities for you to change how you work.

Table 8.2.1 Continued.

Self + Manager	[Self Text +] Additionally, your manager will be able to view the assessments at the end of the week.	The insights from PSAI will be first available to you. By viewing your insights, you can learn more about how you work. This new understanding can identify opportunities for you to change how you work. After some time, your manager will also able to view the insights about you. You will be able to share your interpretation of insights with each other and collaboratively decide approaches to work. The manager will NOT be able to see the data PSAI used to produce insights.
Self + Trusted Other	[Self Text +] Additionally, a coworker of your choice will be able to view the assessments at the end of the week.	The insights from PSAI will be first available to you. By viewing your insights, you can learn more about how you work. This new understanding can identify opportunities for you to change how you work. After some time, a trusted other of your choice would also be able to view the insights about you. Your trusted other can be a close colleague, mentor, wellbeing officer, or anyone you think can you improve your work experience. You can share your interpretation of insights with each other and collaboratively decide on new approach to work. The trusted other will NOT be able to see the data PSAI used to produce insights.
Self + Aggregate	[Self Text +] Additionally, your assessment will be anonymously aggregated to help users compare their experience and learn collective trends.	The insights from PSAI will be first available to you. By viewing your insights, you can learn more about how you work. This new understanding can identify opportunities for you to change how you work. In addition, this PSAI will anonymously pool your insights with other coworkers who have consented. You can compare your experience with different groups, such as others in the same role or department. PSAI will NOT pool insights if the groups are smaller than 50 people to ensure individual identities protected.

Vignette Experiment

I considered the experimental vignette method because it was infeasible to practically and ethically deploy multiple variations of PSAI at scale on real information worker populations. Each vignette represented a scenario where an instance of PSAI is deployed for information work. Generally speaking, each vignette showed a passive data source (H1) that monitors worker behavior in a specific context (H2) to predict either their performance or mental wellbeing (H3) and shares this insight back to certain stakeholders (H4). Table 8.2.1 shows the descriptions of possible components a PSAI system could have because

	Viva	Humanyze	Occupancy	Screenshots	Browse	CCTV	Webcam	
Viva		50.00%	85.71%	100.00%	100.00%	100.00%	100.00%	84.21%
Humanyze	50.00%		80.00%	60.00%	80.00%	100.00%	100.00%	71.88%
Occupancy	14.29%	20.00%		50.00%	33.33%	100.00%	100.00%	37.50%
Screenshots	0.00%	40.00%	50.00%		33.33%	50.00%	100.00%	37.50%
Browse	0.00%	20.00%	66.67%	66.67%		100.00%	100.00%	36.84%
CCTV	0.00%	0.00%	0.00%	50.00%	0.00%		66.67%	17.65%
Webcam	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%		7.14%

Figure 8.2: Participants in the first study (section 8.1) were asked which PSAI they preferred between a pair of systems. The table shows how often a particular system was preferred. System labels correspond to Table 8.2

of variations in the type of sensor and scope of sensing. Table 8.2.1 shows the two different outputs a PSAI system could generate. Lastly, Table 8.2.1 shows the four different information sharing paradigms that involve a PSAI system. These tables not only list the variations in the vignettes but also the accompanying text used to produce a vignette with that component. Accounting for all possible variations, my hypotheses space involves 48 vignettes [3] and 1 baseline vignette.

Baseline Vignette. Since the vignettes vary across categorical variables, I wanted to identify a stable baseline scenario. For this I referred to the participants preferences for different systems as per the interview study (Table 8.2). In that study, each participant was shown pairs of PSAI and asked which technology they were more likely to accept. Participants were shown at least two pairs of PSAI. These two pairs were randomly selected without replacement from seven possible systems. Some participants also evaluated a third pair, where they selected a technology between their first two preferences. In total 28 information workers performed 60 different comparisons. As shown in Figure 8.2, I found that *Sys 7*, based on *Viva Insights* [407], was preferred most often. On the contrary, *Sys 2*, based on *RemoteDesk* [403] was never preferred over another system (table). That PSAI scenario described a webcam analyzing facial expressions and surroundings to measure

[3] $(H1 = 3) \times (H2 = 2) \times (H3 = 2) \times (H4 = 4)$

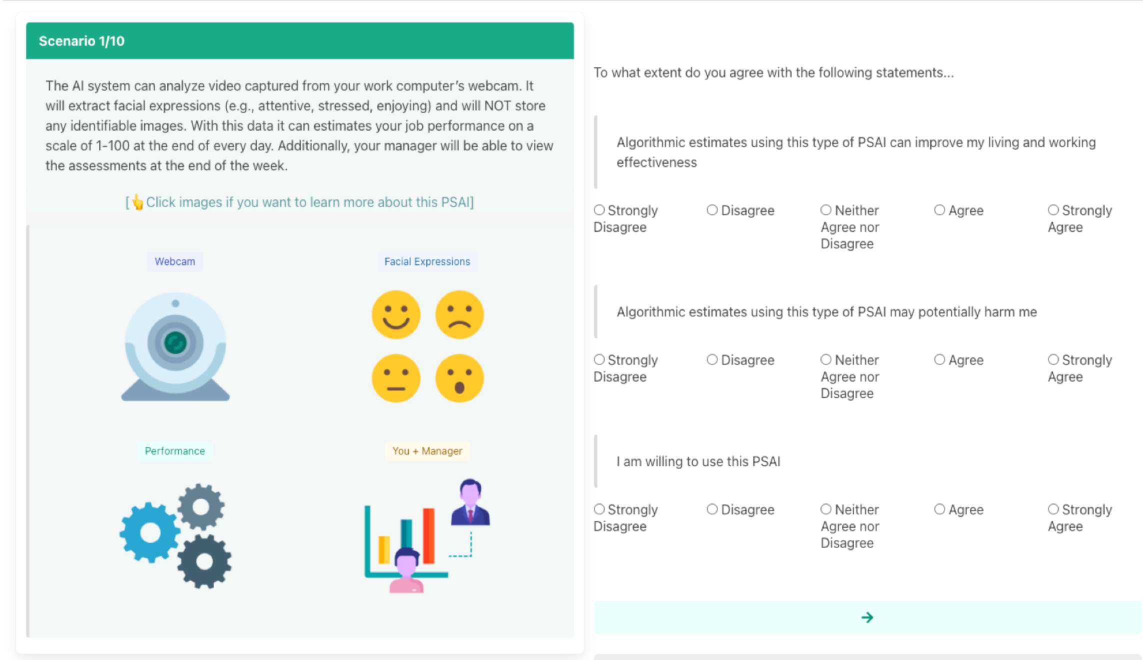


Figure 8.3: The baseline vignette was shown to all participants as the first scenario. For each vignette, participants had to report their perceived utility, perceived harmfulness, and perceived willingness to use the PSAI.

performance and share it with a worker's manager. I constructed my baseline vignette to resemble the characteristics of this technology (Figure 8.3).

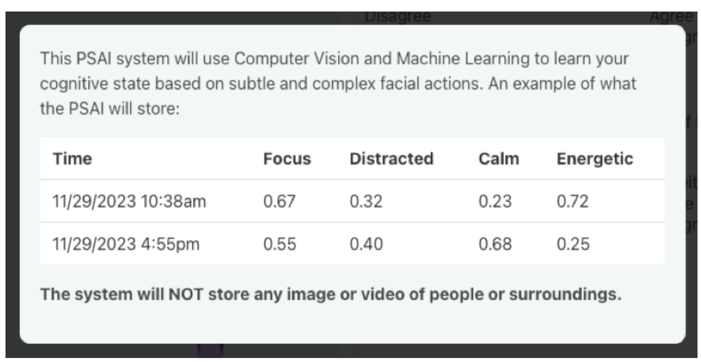


Figure 8.4: Every vignette had 4 icons representing the PSAI in terms of its input, process, output, and users. Participants could click any of these to get a deeper understanding. The example above shows the explanation for *process* in the baseline vignette.

Vignette Evaluation. Every participant assessed a deck of 10 vignettes through an online browser portal. The first vignette for each participant was the baseline. The other 9 were randomly generated combinations from the hypotheses space. The vignettes were presented as a combination of text descriptions along with graphical icons. The icons helped improve recognition of scenarios and emphasize the differences. For any PSAI, the icons corresponded to the input, process, output, and users. Participants could click the graphical icons to learn more in-depth about a specific component of a PSAI vignette. The portal recorded the number of times an in-depth explanation was shown. Before beginning the exercise, participants were given a tutorial of the interface.

For each vignette, participants reported their perceived use, harm, and will to actually use the PSAI in question. In total, they answered 3 questions on a 5–point Likert scale for each vignette. These questions were adapted from Sun *et al.*'s research on location tracking services and privacy caluclus [442]. Specifically, they needed to answer, ”to what extend do you agree with the following statements...”:

1. Utilitarian Benefit: ”Algorithmic estimates using this type of PSAI can improve my living and working effectiveness”

2. Anticipated Harm: ”Algorithmic estimates using this type of PSAI may potentially harm me”

3. Willingness to Accept: ”I am willing to use this PSAI”

I followed many of the practices suggested by Sheringham *et al.* to design a reliable vignette study [443]:

1. *Credibility*: I chose practical factors that information workers considered in the exploratory study to make the vignettes believable.

2. *Number*: Every participant evaluated multiple vignettes to account for individual variances.

3. *Variability*: Each factor occurred in variety of combinations with other factors to represent every possible scenario.

Individual traits and Open-ended responses. After the participants had evaluated their deck of vignettes, they completed an additional questionnaire to report their individual characteristics. This included demographic characteristics (age, gender, race) and the nature of their job (role, size of company, number of employees reporting to them). Abraham *et al.* found that worker's attitude towards *quantification* and *public surveillance* can explain their preference for monitoring at work [401]. I adapted their questionnaires to include a 5–item survey for participants to describe their familiarity with personal tracking technologies and a 3–item survey to express their opinion on public surveillance. To account for additional covariates, I also added a 2–item survey to capture the participants' digital *privacy behaviors* (adapted from [444]). Lastly, Kim *et al.* found that *trust* was a key antecedent in the cost-benefit calculus of such technologies [438]. To accommodate this, I also included a 2-time survey to capture participant trust in their manager (adapted from [445]). Lastly, participants were could answer up to 3 open-ended questions with free-form responses to discuss how they envision PSAI can improve their work, the situations of harm, and designs that protect their best interest.

Participants

The portal launched for public access in March 2023. I advertised the portal through different employee mailing lists, work related social media (e.g., LinkedIn and Reddit), and at physical office spaces. Every exercise session was anonymous and no personally sensitive information was tracked. Before accessing the vignettes, every visitor needed to complete a screening survey. This was the same criteria described in Study 1 along with additional checks for bots. In total, 110 different information workers attempted the vignetted exercise. Karren and Barringer's review found that most vignettes studies involving workers recruited between 80 and 140 participants for a similar vignette space. Collectively, my portal received 1059 evaluations for PSAI vignettes [4]. 90% of the vignettes were evaluated

[4] 9 participants did not evaluate all vignettes in their deck

within 75 seconds, while the median completion evaluation time was about 25 seconds. Since responses might be unreliable due to speeding, I removed 9 vignettes that were completed in less than 5 seconds. Each vignette evaluation was used as a data point in my mixed–effects model to test my hypotheses.

Mixed–Effects Model

I primarily built two *Linear Mixed-Effects* models to understand the impact of various factors on utility and harm. I will refer to these as M_U and M_H. At a high level, every model tested the main-effect of 4 independent variables; $H1 =$Type of Sensing, $H2 =$Scope of Sensing, $H3 =$Type of Insight, and $H4 =$Sharing of Insight. To account for comprehensibility of the vignette, I included the count of in-depth explanations viewed by the worker; E_{input}, $E_{process}$, E_{output}, and E_{users}. And finally, to control for individual and organizational factors, I included covariates (subsubsection 8.2.1). These included demographic characteristics (age, gender, race), job characteristics (role, size of company, number of employees reporting to them), attitude to *quantification*, attitude to *public surveillance*, digital *privacy behaviors* and *trust*. The models fundamentally varied in the dependent variable that I studied. For example, $Y =$Perceived Use in M_U and $Y =$Perceived Utility in M_H. Every participant could evaluate multiple vignettes. Since each evaluation counted as an observation, the model needed to group these together to account for within–participant variances. Therefore, I included the participant as a *random effect*.

$$
\begin{aligned}
Y \sim H1 + H2 + H3 + H4 + E_{input} + E_{process} + E_{output} + E_{users} \\
+ Age + Gender + Race + Org.\_Role + Org._Size + Num\_Reportees \\
+ Perc.\_Quantification + Perc.\_Surveillance + Privacy\_Behaviors + Trust \\
+ 1|Participant \\
Y \in \{perceived\, utility,\ perceived\, harm\}
\end{aligned}
\tag{8.1}
$$

The key variables of interest in these models are categorical (H1, H2, H3, H4). For all the models described in this section I used the *lme4* package in *R* to apply the *lmer*

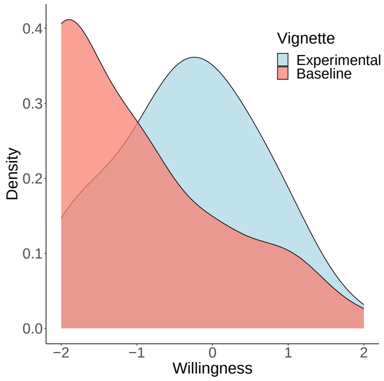
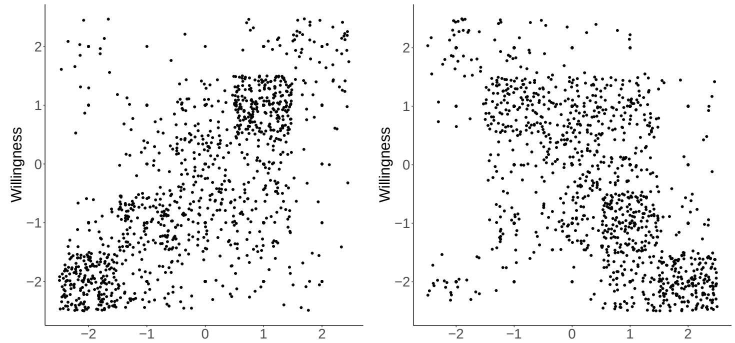


Figure 8.5: The density plot compares the willingness to accept PSAI for each of the 110 sessions. Most respondents were resistant to adopting the system discussed in the baseline vignette.



(a) Perceived Use (b) Perceived Harm

Figure 8.6: The responses received by the Experimental Vignette portal conform with the expected trends from TAM [436]. (a) Increase in perceived use leads to greater willingess to accept PSAI. (b) Increase in perceived harm leads to lower willingess.

function [447]. The *lmer* function creates "dummy" variables for each category internally. In the words of the package's author, De Boeck *et al.*, "the first item functions as the reference item, and that all other item parameters are estimated as deviations from the first" [447]. The reference category for each of the hypotheses was selected to be the same as the baseline vignette — $H1$ =*"visual"*, $H2$ =*"general"*, $H3$ =*"performance"*, and $H4$ ="you+manager".

8.2.2 Findings

Before reporting the findings from the models discussed above, I wanted to confirm some of the assumptions of the experimental design. First, I checked if my baseline vignette was a less acceptable scenario of using PSAI. Table 8.5 shows a strong negative tendency to accept a PSAI that leverages the camera on a worker's personal device to predict performance and eventually share insights with a manager. The distribution validated that the choice of baseline was reasonable. Additionally, I found that sessions showed a normal distribution around 0 for other vignettes. This pattern tells us that workers in the sample are not biased towards accepting or rejecting PSAI and represent a healthy balance.

Second, I checked if the modified version of TAM is an appropriate framework. I built a simple mixed effects model to understand willingness to accept different PSAI (Equation 8.2). The conditional R^2 of this model was 0.72 indicating that a large portion of the variance in acceptability can be explained by the model. Moreover, the relationship between variables was expected. Willingness to accept significantly increased with utility (0.55) and significantly reduced with increased perceptions of harm (-0.42). These preliminary results provided confirmatory evidence that the modified framework to understand the antecedents of utility (M_U) and harm (M_H) can help indicate acceptability of PSAI for information workers (Table 8.6).

$$Willingness \sim Perceived\,Utility + Perceived\,Harm + 1|Participant \tag{8.2}$$

Both my primarily models explained a sizeable portion of variance in their respective target variables (for M_U the $R^2 = 0.60$ and for M_H the $R^2 = 0.58$). These values are improvements over similar studies to explain worker acceptance of technology [438]. Note, however, my model included both fixed and random–effects. On closer inspection, I found that the fixed-effects explain a smaller portion of the variance alone. For comparison, for M_U the marginal-$R^2 = 0.15$ and for M_H the marginal-$R^2 = 0.20$. These measures imply that the unmeasured individual differences (included as the random effect $1|participant$) played a larger role in interpreting perceived use of PSAI than interpreting perceived harms. Having said that, both fixed and random effects were essential to the models. To reemphasize, my hypotheses testing was concerned with the fixed effects, i.e., $H1$, $H2$, $H3$, and $H4$. The following subsections will explain the results of Table 8.7 [5].

[5] Other covariates included in the model were omitted from the table for brevity. These variables contributed to the explained variance and controlled the effects of other IVs. However, in themselves, their estimates itself were not of interest.

Table 8.7: Linear Mixed-Effects Regression models provide insight into the relationship between different variations in PSAI and worker perceptions. The ▲symbol indicates a significant increase when the corresponding component is included in the PSAI. Similarly, ▼indicates a significant reduction. By observing the values in this table, we can estimate which components lead to increased utility (M_U) and reduced harm (M_H), which lead to increased acceptability.
(‘-’:$p < 1$, ‘$\circ$’:$p < 0.1$, ‘*’:$p < 0.05$, ‘**’:$p < 0.01$, ‘***’:$p < 0.001$)

		M_U (Utility)		M_H (Harm)	
		Est.	$p-$value	Est.	$p-$value
H1: Type of Sensing (ref: Visual)	Digital Time Use	0.40 ▲	0.001**	-0.33 ▼	0.002**
	Online Language	0.06	0.589	-0.16	0.145
	Physical Activity	0.40 ▲	0.001**	-0.48 ▼	9×10^{-6}***
H2: Scope of Sensing (ref: General)	Work (only)	0.05	0.351	-0.04	0.44
H3: Type of Insight (ref: Performance)	Mental Wellbeing	0.14 ▲	0.021*	-0.15 ▼	0.004**
H4: Sharing of Insight (ref: Self + Manager)	Self (only)	0.54 ▲	4×10^{-10}***	-0.53 ▼	2×10^{-11}***
	Self + Aggregate	0.18 ▲	0.031*	-0.34 ▼	3×10^{-5}***
	Self + Trusted Other	0.08	0.346	-0.12	0.10 $\circ$
Explanations	Input	0.02	0.834	-0.02	0.841
	Process	-0.02	0.856	-0.17 ▼	0.09 $\circ$
	Output	-0.04	0.732	0.12	0.278
	Users	0.01	0.872	-0.03	0.709

H1: Physical Activity and Digital Time Use are more acceptable

Many of the PSAI scenarios presented in my vignette experiment exist as alternatives to the aggressive surveillance of commercial options. A common form is encapsulated by *RemoteDesk* [403]. In such technologies, the camera is leveraged as the source of sensing a worker's behaviors. The type of sensing is “visual”. My results show that in comparison to recording a visual stream of data, other types of sensing are more favorable. Table 8.7

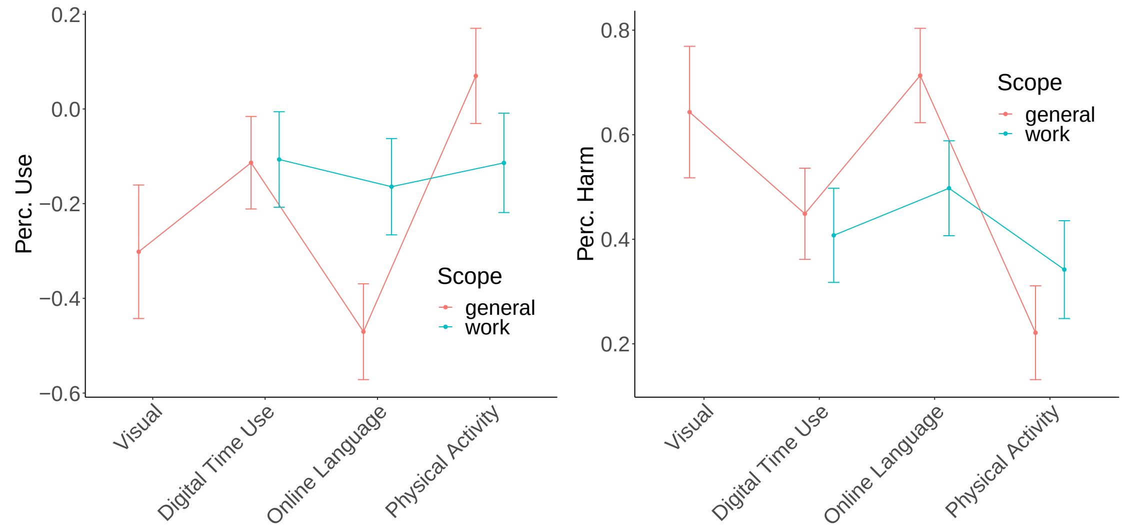
(M_U) shows a significant positive coefficient for recording digital time use of applications (estimate= 0.401, p-val= 0.001) and for tracking physical activity patterns with wearables and or embedded devices (estimate= 0.401, p-val= 0.001). However, the results indicate little additional utility for mining online language from communication tools and social media. Based on these results I rejected the null hypotheses that perceived usefullness of PSAI is independent of the type of sensing. Thus, hypotheses *H1a* holds.

The relationship of these variables with harm (M_H) are symmetrical. The perceived harm for PSAI reduced when it was sensing digital time use (estimate= -0.33, p-val= 0.002) or physical activity (estimate= -0.48, p-val= 7×10^{-6}). It is also worth noting that, PSAI systems that use of online language can be perceived to be less harmful than monitoring with a camera (estimate= -0.16, p-val= 0.145), but the effect is not statistically significant. Together, type of sensing is related to the perceived harm of different technologies. Therefore, hypotheses *H1b* is likely to be true.

These results also shed more light on the preferences from the exploratory study (Figure 8.2). and were more preferred. The most preferred system, *Sys7*, was inspired by *Viva Insights* [407]. This application primarily records the time and event counts of digital activities such as communication and document use. The next in line was *Sys5*, which was inspired by *Humaneyze* [389]. Interestingly, *Humaneyze* provides multimodal approach which is a mix of digital time use, online language, and physical activity. In chapter 4, I had shown the value of combining multiple streams to provide a full picture. These findings suggest that inclusion of certain sensors can increase resistance to adopting PSAI, even if it is more likely to provide a fuller picture of the worker.

H2: Work–Life Scope only matters in conjunction with Sensing Stream

Presently, it is common to work remotely from one's home or from a different spot away from a designated office space (e.g., a coffee shop). Even though some workers have embraced this spatial flexibility, it raises concerns about the limits of sensing. This concern



(a) (b)

Figure 8.7: Scope of Sensing (*H2*) interacts with Type of Sensing (*H1*) when trying to model (a) utility and (b) harm. Although *H2* itself did not have a significant main effect in M_U and M_H, it is evident in the case of online language that the general scope has lower utility and higher harm than work scope.

is similar to the finding boundaries where work stops. M_U and M_H included variables to compare work–only scope of PSAI with a broader scope, where a worker's activities outside of work are also sensed. I found that this distinction did not significantly reflect any changes in the perception. Therefore, the null hypothesis still holds, i.e., scope of sensing does not indicate the utility or the harm of PSAI at work.

Different workers have different perspectives on work–life segmentation. It is possible that the models could not converge at a general solution. Also note, that unlike other components, the scope of sensing was tied to the type of sensing in how it manifested in each vignette (Table 8.2.1). Therefore, to get a better understanding of the scope of sensing, I ran a post-hoc regression analyses where *H2* interacts with *H1* (Equation 8.3).

$$
\begin{gathered}
Y \sim H1 + H2 + H3 + H4 + H1 : H2 \\
Y \in \{perceived\ utility,\ perceived\ harm\}
\end{gathered}
\tag{8.3}
$$

I found that the work scope interacts with the sensing online language when it comes to both perceived utility and harm. Table 8.8a and Table 8.8b show the interaction plots between the two variables based on the effects fitted by the interaction models. It is evident that online language is considered more useful and less harmful when it is constrained to work-specific applications and platforms. In fact, the broad sensing of online language is possibly worse than the baseline vignette. Studying the interaction also confirms that the scoping does not play a significant role in describing the acceptability of other streams.

H3: Mental Wellbeing Insights are more acceptable

Most of the commercial systems in use today are focused on providing performance insights. This trend possibly stems from the organizational need to maintain productivity. Looking at the workers' perspective tells us a different story. Vignettes where PSAI provided insight on stress were not only considered less harmful (estimate= -0.15, p-val= 0.006), but also more useful (estimate= 0.14, p-val= 0.021) than those that estimated performance. Based on these results, I argue that both *H3a* and *H3b* are true.

In a workplace, measures of performance are directly linked with extant evaluation metrics that could eventually determine promotions or layoffs. In contrast, a worker's mental wellbeing often needs more personal management and organizations are only starting to support worker mental wellbeing. Among today's enterprise technologies, *Viva Insights* [407] stands out because it provides mental wellbeing insight in addition to performance insights. Today's workplace has a dearth of PSAI systems that specialize in supporting worker mental wellbeing. My findings motivate the development of PSAI to algorithmically infer constructs that workers are interested in, such as stress.

H4: Keeping insights private or sharing as aggregate is more acceptable

In the previous study, workers were able to appreciate the need to share insights but they needed the insights to first be provided to them. Then they get to choose when and how it is shared forward. The baseline vignette depicted sharing with the manager after it had been sent to the worker. Although the manager might be able to supervise tasks better, the regression analyses shows that workers found it significantly more useful to keep the insights to themselves (estimate= 0.54, p-val= 4×10^{-10}). They also found value in sharing their insights as an aggregate for collective interpretation (estimate= 0.18, p-val= 0.031). Keeping the insights private (estimate= 0.53, p-val= 2×10^{-11}) or sharing it as an aggregate (estimate= -0.34, p-val= 3×10^{-5}) were considered significantly less harmful. Sharing the insights to trusted others, such as peers, seniors, or mentors, did not have more utility. However, it was considered less harmful when considering a more liberal confidence interval (estimate= -0.12, p-val= 0.10). Taken together, I rejected the null hypotheses that PSAI are perceived useful and harmful regardless of who the information is shared with. As a result, both *H4a* and *H4b* hold true.

In Study 1 six of the seven systems sent insights about the worker to a different stakeholder. The vignette experiment provides evidence that sharing individualized insights with managers (or immediate supervisors) is not a generally accepted approach. Instead, we need to consider the personalized designs of PSAI that give the locus of control to the worker. Arguably, this signals a shift in how these technologies are perceived and deployed today. Organizations can still benefit from the personal approach by focusing on aggregate views of the PSAI outputs to evaluate collective trends without singleing out individual workers. Additionally, sharing to worker insights to specific trusted individuals might be safe to explore in contexts such as counseling.

Robustness Analyses

Beyond the hypotheses tests, I wanted to further probe the validity of varying information flows of PSAI. This subsection describes additional findings from my experimental vignette study that reinforce the value of carefully designing PSAI.

Perceived utility and harm can mediate the effect of PSAI components on acceptability. The fundamental idea of TAM states that the willingness to adopt technology is a function of increasing utility and lowering costs [436]. My empirical data supports this idea by showing how perceived utility and perceived harm can describe the willingness to adopt a given PSAI system (Table 8.6). Additionally, through M_U and M_H I have presented the various aspects of PSAI that impact how it is perceived by workers. These results motivate me to hypothesize that certain components of PSAI — as represented by ***H1, H2, H3, H4*** — are likely to result in more acceptable designs of PSAI. Note, however, neither of those regression models could completely account for the variances observed. Perceived harm could be influenced by certain unmeasured factors that were external to my experimental design. In theory, it is possible that increased perception of harm through such exogenous aspects subdues the value of certain sensors or sharing paradigms. Therefore, I decided to validate if selection of certain components actually impacts the willingness to accept PSAI.

$$
\begin{aligned}
Willingness \sim H1 + H2 + H3 + H4 + \\
+ < Same\ covariates\ as\ M_U\ and\ M_H > \\
+ Mediator \\
+ 1|Participant \\
Mediator \in \{perceived\ utility,\ perceived\ harm\}
\end{aligned}
\tag{8.4}
$$

To disentangle this, I conducted *causal mediation analyses* [448]. The aim of this exercise was to determine how modifying PSAI can directly or indirectly effect acceptability. Findings from Table 8.7 already establish the relationship between PSAI components and the mediators. Additionally, M_W discussed in the preliminary findings also confirms the relationship between the mediators and willingness. To test mediation, I fit regression models to explain willingness to accept with the PSAI components and the mediator (Equation 8.4)— $M_W U$ and $M_W H$ corresponding to the different mediators. Then, I tested for the effects using the *mediation* package in *R* [449]. Since this approach only accounts for binary independent variables, I reported findings only for specific cases that supported the hypotheses. Table 8.9 shows that the *average causal mediation effect* (ACME) was significant and positive for type of sensing, type of insight, and sharing of insight. These values

Table 8.9: Causal mediation analyses help confirm the relationship between PSAI variables, perceptions of the technology, and willingness to accept it. ACME denotes the average causal mediation effect, which is the indirect impact of the treatment on acceptability. ADE denotes average direct effect, which indicates the direct impacts that cannot be explained by the mediator.
This table does not report raw p-values for the sake of brevity. Significant effects have been labelled with the following scheme: '-':$p < 1$, '$\circ$':$p < 0.1$, '*':$p < 0.05$, '**':$p < 0.01$, '***':$p < 0.001$

		Mediator: Utility		Mediator: Harm	
	Treatment	ACME	ADE	ACME	ADE
Type of Sensing (control: Visual)	Digital Time Use	▲0.19**	▲0.2*	▲0.25***	▲0.25 *
Type of Insight (control: Performance)	Mental Wellbeing	▲0.09*	0.05	▲0.08**	0.05
Sharing of Insight (control: Self + Manager)	Self (only)	▲0.35***	▲0.42***	▲0.30***	▲0.47***

imply that making changes to PSAI information flow can lead to increased acceptability because of their affect on perceived utility and harm.

Learning more about feature processing can reduce the perceived harm of PSAI Whenever a participant viewed a vignette, they could click each of the icons to learn more about certain aspects of that PSAI system. The portal recorded these explanation–related interactions for approximately 23% of the vignettes. Studies show that adding transparency to the AI–based black boxes can have its benefits [450]. The results of my mixed-effects regression models (Table 8.7) showed that when workers saw explanations of the underlying process, they perceived less harm. An example of the explanation for process is shown in Figure 8.4. Not only does it describe the types of features extracted from the sensor source, but it also explicitly states common kinds of artifacts that will not be recorded. Note, however, this experiment was not designed to study explanations. Also, the models capture explanation seeking behaviors, not the fidelity of the explanations itself. Therefore, these results are not conclusive, but do motivate additional experimentation to study the value of

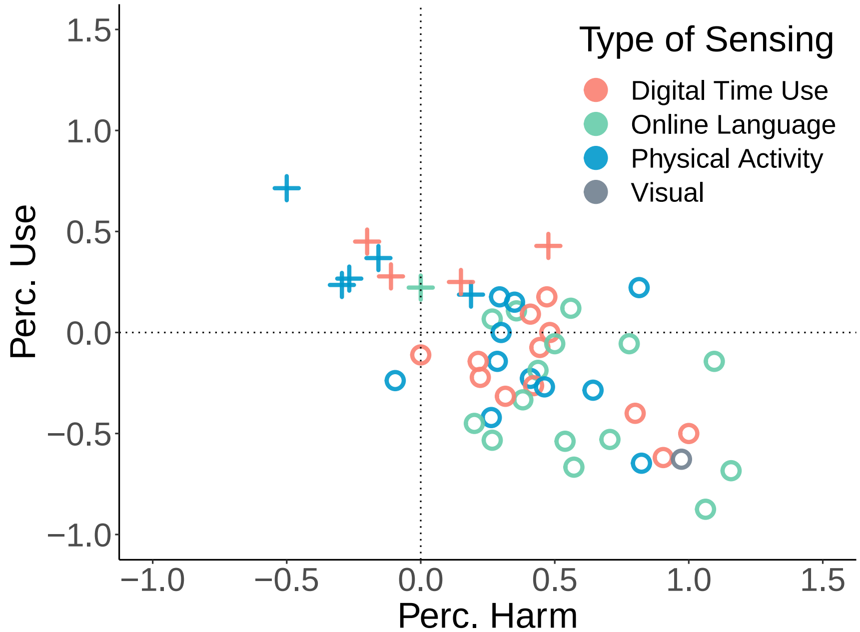


Figure 8.8: Each point on this plot represents 1 of the 49 PSAI vignettes represented in my experiment. + represents vignettes had average acceptability greater than 0.

explainability in PSAI.

Acceptable designs of PSAI are limited. The nature of vignette experiments makes the comparisons relative. In this study, the baseline vignette was carefully selected as a reference. Camera based sensing for performance inferences was poorly received in Study 1. Results from this study confirmed the low willingness to adapt as well (Table 8.5). With that reference point in mind, the regression models indicate that improvements can be made in how PSAI is deployed. The question is, if these improvements are enough? The experimental setup covered 48 possible scenarios where a worker interacts with PSAI. The 110 different participants evaluated many of the same vignettes to indicate their willingness to use that particular instance of PSAI. The score ranged from -2 to $+2$, where a negative score indicated resistance to adoption. Figure 8.8 shows that only 10 of the 49 vignettes (including the baseline) had a positive average for willingness score. Therefore, a large proportion of the possible implementations of PSAI are less likely to be worker-centric. While each vignette was a combination of components, the participants evaluated each of these as an integrated scenario where the moving parts were not explicit. Careful inspecting actual scores reveals that even performance–based insights are acceptable in the right combination (Table 8.2.2). For example, in vignette #33, the PSAI system infers performance by sensing physical activity. Interestingly enough, vignette #22 and #23 were both worse than the baseline. From the findings, it is clear that this negative reaction is a combination of sensing online language in broad social media to predict performance and then eventually share it beyond the data subject. As it stands right now, without considering worker needs most implementations will fail and only exacerbate poor work experiences. Referring to these raw scores before implementing PSAI information flows can help anticipate their adoption (or resistance).

Table 8.10: Each vignette was evaluated by several different information workers as a part of the experiment. This table shows the average scores for Perceived Utility, Perceived Harm, and Willingness to Accept. The vignettes are sorted in decreasing order of acceptability. Vignette #49 was the baseline vignette shown to all participants.

ID	H1	H2	H3	H4	Utility	Harm	Accept
33	Physical Activity	General	Mental Wellbeing	Self	0.71	-0.50	0.64
34	Physical Activity	General	Mental Wellbeing	Self + Aggregate	0.27	-0.27	0.40
13	Digital Time Use	Work	Performance	Self	0.28	-0.11	0.33
37	Physical Activity	General	Performance	Self	0.19	0.19	0.25
9	Digital Time Use	Work	Mental Wellbeing	Self	0.43	0.48	0.19
10	Digital Time Use	Work	Mental Wellbeing	Self + Aggregate	0.25	0.15	0.15
1	Digital Time Use	General	Mental Wellbeing	Self	0.45	-0.20	0.10
44	Physical Activity	Work	Mental Wellbeing	Self + Trusted Other	0.24	-0.29	0.06
26	Online Language	Work	Mental Wellbeing	Self + Aggregate	0.22	0.00	0.06
41	Physical Activity	Work	Mental Wellbeing	Self	0.37	-0.16	0.05
29	Online Language	Work	Performance	Self	0.11	0.36	-0.04
38	Physical Activity	General	Performance	Self + Aggregate	-0.24	-0.10	-0.05

Table 8.2.2 Continued.

14	Digital Time Use	Work	Performance	Self + Aggregate	-0.11	0.00	-0.06
46	Physical Activity	Work	Performance	Self + Aggregate	-0.23	0.41	-0.09
21	Online Language	General	Performance	Self	-0.53	0.27	-0.13
25	Online Language	Work	Mental Wellbeing	Self	0.07	0.27	-0.13
2	Digital Time Use	General	Mental Wellbeing	Self + Aggregate	-0.07	0.44	-0.15
5	Digital Time Use	General	Performance	Self	0.00	0.48	-0.15
42	Physical Activity	Work	Mental Wellbeing	Self + Aggregate	0.15	0.35	-0.15
17	Online Language	General	Mental Wellbeing	Self	0.12	0.56	-0.16
28	Online Language	Work	Mental Wellbeing	Self + Trusted Other	-0.06	0.78	-0.22
36	Physical Activity	General	Mental Wellbeing	Self + Trusted Other	-0.14	0.29	-0.24
6	Digital Time Use	General	Performance	Self + Aggregate	-0.22	0.22	-0.28
45	Physical Activity	Work	Performance	Self	-0.29	0.64	-0.29
31	Online Language	Work	Performance	Self + Manager	-0.19	0.44	-0.31
30	Online Language	Work	Performance	Self + Aggregate	-0.54	0.54	-0.35
15	Digital Time Use	Work	Performance	Self + Manager	-0.26	0.42	-0.37
48	Physical Activity	Work	Performance	Self + Trusted Other	-0.42	0.26	-0.37
12	Digital Time Use	Work	Mental Wellbeing	Self + Trusted Other	-0.14	0.21	-0.39
40	Physical Activity	General	Performance	Self + Trusted Other	0.18	0.29	-0.41
20	Online Language	General	Mental Wellbeing	Self + Trusted Other	-0.33	0.38	-0.48
8	Digital Time Use	General	Performance	Self + Trusted Other	0.09	0.41	-0.50
39	Physical Activity	General	Performance	Self + Manager	-0.27	0.46	-0.50
3	Digital Time Use	General	Mental Wellbeing	Self + Manager	0.18	0.47	-0.59
18	Online Language	General	Mental Wellbeing	Self + Aggregate	-0.45	0.20	-0.60
43	Physical Activity	Work	Mental Wellbeing	Self + Manager	0.00	0.30	-0.60

Table 8.2.2 Continued.

24	Online Language	General	Performance	Self + Trusted Other	-0.14	1.10	-0.62
32	Online Language	Work	Performance	Self + Trusted Other	-0.06	0.50	-0.67
35	Physical Activity	General	Mental Wellbeing	Self + Manager	0.22	0.81	-0.67
11	Digital Time Use	Work	Mental Wellbeing	Self + Manager	-0.50	1.00	-0.70
7	Digital Time Use	General	Performance	Self + Manager	-0.62	0.90	-0.71
19	Online Language	General	Mental Wellbeing	Self + Manager	-0.67	0.57	-0.71
27	Online Language	Work	Mental Wellbeing	Self + Manager	-0.53	0.71	-0.76
4	Digital Time Use	General	Mental Wellbeing	Self + Trusted Other	-0.32	0.32	-0.79
47	Physical Activity	Work	Performance	Self + Manager	-0.65	0.82	-0.82
16	Digital Time Use	Work	Performance	Self + Trusted Other	-0.40	0.80	-0.93
49	Visual	General	Performance	Self + Manager	-0.63	0.97	-1.06
22	Online Language	General	Performance	Self + Aggregate	-0.68	1.16	-1.11
23	Online Language	General	Performance	Self + Manager	-0.88	1.06	-1.38

8.2.3 Summary

I have shown that different types of sensors impact perceptions of utility and harm for PSAI differently (*H1*). Restricting sensing to the work scope can be favorable for certain sensor streams (*H2*). Developing PSAI to infer mental wellbeing insights was better perceived than those that measuring performance (*H3*). Lastly, workers more value in keeping PSAI–generated insights private or shared as aggregate than forwarding it to specific individuals in their organization. These sharing paradigms also assuaged their concerns of harm (*H4*).

CHAPTER 9

DISCUSSION: SHAPING THE FUTURE OF WORK

In 1930, Keynes described a future economy where we would work fewer hours than we actually do today[451]. His prediction underestimated many socio-economic factors, but it also misrepresented what today's workers might want in terms of purpose, self-worth, and leisure [452]. Information workers today do not want to work more, they want to work better. For some, being successful is about performing to their full capacity. For others, it is meeting their role's requirements with better control of their stress and anxiety. Yet, this has been especially challenging to achieve in practice because organizations are not taking care of their workers' needs. This phenomenon is exemplified by the social media trend known as *Quiet Quitting*, where information workers showed an increasing tendency to detach from work while remaining employed [453]. The rise of unhappy workers indicate that we are facing an *emotional recession* [454]. Although Keynes' vision was inaccurate, he highlighted that the future of work needs to be centered on improving the lives of workers. Recent literature suggests that we can thrive as workers by adopting healthier work practices [455]. However, organizational sciences have had limited personalized insight into the effectiveness of information workers because of relying on traditional survey–based approaches [4, 5]. My thesis posits an alternative method to help workers gain insight into healthier work contexts. Computing technologies have played a central role in an information worker's relationship to work and even their life outside of it. Workers are constantly leaving traces of their activities on their personal computers, smartphones, social media, and other networked devices embedded in their surroundings. This dissertation demonstrates that same technology that information workers interact with daily can be repurposed to explain their performance and mental wellbeing and develop personal informatics applications for workers to succeed.

The early chapters in my dissertation have showed the technological feasibility of recording day–level behavioral dynamics and applying computational methods like machine learning can give us a new perspective on identifying effective workers. Chapter 4 presented how multimodal sensing of worker activities can complement personality assessments of workers to provide a more holistic understanding of performance, which also offered workers an opportunity to change. Chapter 5 demonstrated how everyday technologies can model routines of workers to produce a new behavior–based measure to re-organize workers into teams they will thrive in the most. Chapter 6 showed that anonymous, aggregated, and archived data revealed longitudinal patterns in organization behavior that could then help make better organization–level decisions to support worker wellbeing. Through these chapters I illustrated how everyday digital technology can be leveraged for passively sensing workers' ecology at the individual, team, and organizational level. In the subsequent chapters, I studied the challenges we will face in using these methods to produce new applications for workers. Chapter 7 provided evidence for methological constraints in building prediction applications for worker wellbeing and also demonstrated approaches to overcome these constraints. Finally, Chapter 8 investigated the workers' perspective on Passive Sensing–enabled AI(PSAI) and how it can be instantiated to both help and harm workers from meeting their wellbeing needs. Taken together, these studies provide a comprehensive view of passive sensing for worker wellbeing. On one end, this research promotes the idea of using passive sensing to gain dynamic insights into information worker wellbeing by considering various ecological factors. On the other end, this research defines the practical boundaries of making meaningful applications for workers that will support their aspirations of working better.

It is worth noting that the overarching view provided by my studies is a view through the lens of an applied computer scientist. I took a pragmatic approach in improving information work as we know it today. The implications are not ”best practices” but rather ”better practices” — alternatives to methods before it. Yet, we need to be wary of how passive sensing interacts within the complex socio-economic ecosystem of work. A true dent in the lives of workers needs many changes beyond the scope of this dissertation. Beyond technological improvement, we need reform in our research methods and policy. Through this discussion chapter I present a starting point, from where future researchers can reflect and rethink how wellbeing informatics can make working thrive.

9.1 Need for a Worker-Centered Perspective

One of the core motivations behind passive sensing applications is the opportunity to study human wellbeing in natural environments. In today's urban environments, people are likely to be equipped with a variety of connected devices (e.g., smartwatches), they regularly engage with digital platforms (e.g., social media), and are exposed to other embedded technologies in their surroundings (e.g., bluetooth beacons). Each of these modalities can provide valuable behavioral traces. Before the ubiquity of these sensing streams in our daily life, researchers attempted to instrument fixed structures to understand free living behaviors. The *Aware Home* at Georgia Tech [456] and the *PlaceLab* at MIT [457] are examples of such instrumentation. In the twenty years since, personnel management have considered similar ideas to build "smart offices" to support thriving workers [458]. Arguably, surmounting the engineering challenges of such endeavors is non–trivial. However, these projects need to think beyond technological efficacy and also about the perspectives of its data providers, the information workers. In this section, I am going to use one such recent project, *Mites* at CMU [459], as an example to illustrate why we must align sensor deployments with worker perceptions of adoption.

9.1.1 Tensions in Deploying Passive Sensing in Work Contexts

The *Mites* project represents a "unified, high-fidelity, and general-purpose sensing system" for smart buildings [459]. To a large extent, the goal of this project was to better utilize infrastructural resources for building sustainability and occupant wellbeing. Having said that, human activity modeling was one of the proposed applications, which aligns with the kind of individualized insights for workers I have discussed in my studies. Another core difference was that *Mites* did not leverage existing connected devices but rather innovated on a singular, miniaturized, multimodal sensor. The research team fitted 334 sensors across the spaces in one of the campus buildings. If we use *Privacy by Design* for ubiqutious computing as a benchmark [420], the overall sensing framework and the research project was setup with privacy–preserving principles. The sensors were connected through end-to-end encryption in a university–only network (*Adequate Security*. Any data recorded by the sensors was first featurized in the sensor itself, therefore ensuring raw data never left the physical device (*Locality*). The location signatures of these sensors were obfuscated to make re-identificaton of occupants more difficult (*Pseudonymity*). In fact, users were able to opt-out of sensing through a companion mobile application, giving some sense of *Choice and Consent*. Lastly, the research team communicated the role of these sensors through a town-hall, email threads, and placing QR codes (linked to documentation) across different rooms (*Notice*). Despite these safeguards, however, the instrumentation of an existing building into a living laboratory was met with resistance [460]. This pushback stemmed from a gap between the occupants' anxieties, the researchers' intentions, and the technology's capabilities. Eventually 9 of 110 offices disabled the sensor [1]. It can be tempting to consider this small proportion to indicate the success of smart instrumentation of offices. However, I urge future researchers to be careful in associating this case-study as a reason to sense workers indiscriminately.

9.1.2 Foreseeing the Adoption of Passive Sensing in Work Contexts

We learned in section 8.1 that the contextual norms of information workers are unique. Trust runs much thinner in organizational settings than in research environments. Universities tend to be more flexible and heterogeneous. Occupants might not need to visit the

[1]The project was still ongoing when this document was compiled

space instrumented with sensors and become indifferent. For a worker, the danger is that they might not be able to change where they work. Given appropriate mechanisms, the number of opt-outs from such a project would be much higher in an information work setting. In reality, PSAI can be imposed on workers and their inability to freely consent can worsen their overall happiness at work. Let us take another look at *Mites* from the perspective of the components that inform perceptions of PSAI's utility and harm (section 8.2):

1. *Type of Sensing: Mites* possessed 12 kinds of sensing streams. Some of these are beyond the ones I studied, especially the infrastructure–specific streams, such as room temperature, light, humidity, and pressure. It is unclear how these would be perceived. *Mites* included a sophisticated motion sensor to determine *physical activity*, which I found was perceived better than cameras. However, the sensors also included audio sensors which might be perceived as unfavorably as *online language* — a verbal communication stream.

2. *Type of Insight:* Another big challenge with *Mites* was that occupants did not receive clear insight. The sensing suite provides rich possibilities, including insight into occupant stress. However, the lack of actual consumable insights can make it difficult for occupants to envision utility without anticipating the harms of surveillance.

3. *Sharing of Insight:* While the occupants themselves were not receiving any insights, the investigators could gain some insight for benchmarking purposes. I had demonstrated earlier that data providers of PSAI must receive some insight from the data to perceive value and evaluate the harm. If occupants or workers are isolated from the information flow after their data is collected, they are less likely to trust their data will be interpreted accurately.

The assessment above did not need the development of sensors. In fact, it showcases how passive sensing flows can be evaluated *a priori* to ensure worker-centric deployments.

As technologists, we often forget that real world deployments, need us to care beyond specific engineering aspects and consider the entire data flow including critical stakeholders. By doing so, we can ask critical questions regarding stakeholders, data control and benefits.

9.2 Worker-centric Applications through and for PSAI

One of the core ways to make PSAI acceptable technology for workers is directly providing insights back to them (chapter 8). Such a transaction between data providers and a computational tool protects worker interests. Quantified work needs to surface sensing insights in a way that workers have agency to make meaning out of it [461]. Once workers can comprehend their data in the context of inferences, they can choose if they want to involve other stakeholders. Throughout, they should be able to identify the blindspots and misrepresentations of these tools (chapter 7). Not only will this give them a deeper understanding to revoke their involvement in the system but also an opportunity to gather new information required to fully discuss their performance and wellbeing insights with other stakeholders. I believe we need to build prototype interfaces for workers when we begin data collection. Thus, workers will be able to assess PSAI transparently in parallel to the data collection, not after they lose control and possession of their data. These interfaces must follow the following key tenets:

1. *Worker–facing:* Workers must be able to view insights from the data they provide.

2. *Worker–first:* Workers must receive timely and updated insight to help them control the data they provide in the future.

3. *Worker–flexible:* Workers must be able to adjust and contest any algorithmic insights.

In this section I describe some of these potential worker-centric applications that can be inspired from my research. These applications can help workers improve but also help them inspect how they are estimated through PSAI. The interfaces are meant to encourage recourse and engagement from the workers.

9.2.1 Applications for Workers to Evaluate Themselves

We need to conceptualize applications of worker wellbeing as personal informatics tools. In chapter 4, I showed how multimodal sensing could describe *organizational personas* that express workers' performance based on mutable daily activities. Not only were these stylistic representations of the different kinds of workers, but they also revealed testable hypotheses that could be verified with further experimentation. Workers should be able view the daily activities they consent to as factors contributing to their work–related outcomes. This can help them determine if they should change the way they sleep or install screen time management tools. Similarly, we can imagine interfaces where workers evaluate themselves as a function of collaborative teams. Chapter 5 revealed the importance of behavioral views on P–O Fit. Applications that communicate these dynamics can provide workers more agency to reflect and positively influence their performance and wellbeing. When workers are aware of their (lack of) fit, they can carve out their own unique place in a team. This insight can empower them to adopt measures that define boundaries between one's own preferences and the behavioral norms. These interfaces should be longitudinal and provide insights within the context of organizational life cycle. Workers should be able to compare their effectiveness across different organizational groups during times of organizational crises, upheavals, or unanticipated policy changes or enforcement within the organizations. Such interfaces can give workers a robust illustration of themselves but also a means to reflect on how their data can be interpreted.

9.2.2 Applications for Workers to Evaluate Organizations

Workers should be able to leverage their data to keep their organizations accountable. In the days of social media, it is not uncommon to view crowd contributed posts describing companies [241]. These platforms are used by workers to corroborate, compare, and contrast their experiences. It provides a method for job seekers answer questions like "how is it like working in company X?", or "how healthy is our culture?". Chapter 6 provided evidence that free-form language on these platforms can be used to express the organizational culture of different organizations and departments within them. It is not hard to imagine that we build such a knowledge base or "wiki" by accumulating inferences from other kinds of more acceptable passive sensing, such as physical activity or digital time use. With this information, workers could obtain an empirical understanding of how an organization's practices align with their personal values, beliefs, and work ethics. Much like making sense of their own information, aggregating information across workers can help them keep the organization accountable and give them the transparency with which they can investigate the changes in worker wellbeing.

9.3 Directions for Worker-centric PSAI

My studies highlight opportunities through which PSAI can mitigate, but also aggravate, the *power asymmetry* at work. Through this section I aim to inspire socio–technical changes and reflection on how information flows involving PSAI are deployed at work. I envision that these changes need to be centered not only on the development of these technologies but also on structured workplace policies and cultural reforms that encourage a different relationship between workers and their behavioral data.

9.3.1 Align Work–Life Boundary Preferences

The normalization of remote work has made IWs ambiguate which devices are considered work and nonwork, such as the mobile phone [462]. PSAI leverages such devices to model behaviors. In chapter 8, IWs had distinct preferences for the scope of PSAI vis-a-vis their work–life boundary. We need to recognize that IWs have different preferences on how they combine and contrast their private and professional lives. For instance, some workers choose to disclose more of their personal situation to their managers than others who prefer to keep it separate from work (subsubsection 8.1.2). Another way to view this dichotomy is by recognizing different strategies to adjust to the asymmetries at work. More information can lead to more power [463]. IWs who prefer segmentation in PSAI might want to ensure the organization does not get any more power. IWs who want complementary information from PSAI might want to increase the power they have. Arguably, it is challenging to reconcile both perspectives and arrive at a universal PSAI template. Having said that, designers of these systems must be sensitive to these individual preferences when trying to solicit consent and provide notice. Describing the technology on these lines can support more informed decision–making for adoption.

9.3.2 Accommodate Pluralistic Models of Effectiveness

Information work allows workers to approach their work the way they like [426]. I had found that certain IWs anticipated that algorithmic inferences produced by PSAI would take away that discretion. They suspected that such technology will measure all IWs by the same yardstick. It would be unfair to their unique approach to work (subsubsection 8.1.2). This concern could stem from the inability of IWs to determine their own evaluation criteria in an asymmetrical power structure [464]. Hence, IWs resign to accepting that PSAI will also impose—or be used to impose— rigid terms. The domain of organizational psychology already has some precedent for this pluralistic notion. Research on PSAI grounded in such work has incorporated measures beyond task proficiency, like organizational citizenship, to define performance [111, 14]. However, my research on PSAI has also shown that for abstract constructs like wellbeing, algorithmic inferences can be semantically disconnected from what is perceived (chapter 7). The shift to hybrid work has further compelled the need for a more diverse view of effectiveness that might even include domestic activities [465]. Recent research on hybrid information workers showed that technological interventions can support worker wellbeing by helping them meet their diverse set of work–related goals by giving them more agency on their time [466]. Therefore, PSAI systems need to work more intimately with an IW to retrain themselves based on the data subject's uniqueness but also infer insights that speak towards their goals.

9.3.3 Setup Affordances for Human Reappraisal

Generally speaking, IWs did not want to share raw data collected by PSAI, but they still acknowledged that in certain circumstances the estimates output by PSAI needed to be distributed. As learned in subsection 8.1.3, they might want feedback from coworkers or want to compare themselves. The critical challenge in such flows is to protect the IW from being misrepresented by algorithmic estimates. Arguably, PSAI can produce evaluations automatically and at a higher frequency than traditional organizational methods, but these evaluations need to be complemented with human expertise and perspective. Recent studies show that IWs might have very different understandings of their behaviors than what can be measured by PSAI [467]. Conversely, IWs might not be able to interpret personalized insights or conceive actionable changes without the support of their managers or mentors (subsubsection 8.1.3). Both these use cases represent the need for human–in–the–loop information flows that encourage reappraisal by data–subjects and experts. Whom a worker considers an expert might vary from person to person. Rawls believed that impartial experts and mutual accountability can form a social contract that can legitimize social control, such as that being proposed by PSAI [468]. Since HR typically has a contentious reputation among IWs, organizations might need to appoint specialized officers for this role, such as the up-and-coming wellness officers. Although, my findings from subsubsection 8.2.2 showed that IWs are unlikely to adopt systems that automatically include even trusted others. It is concerning that workers today are unable to identify credible and trustworthy human experts to interpret their work wellbeing. Having said that, in personal contexts of passive sensing, we now have some evidence of semi-automated methods of data sharing that respect privacy preferences [469]. Perhaps, the solution lies in stakeholders outside of their work context or other policy reform to protect worker interest.

9.3.4 Design for Worker Undersight

Given the nature of the data it captured, PSAI was viewed as empirical evidence that could drive changes in an IW's professional state but also in their overall organization. If an IW was overworked, they could convince their manager to rearrange work distribution or give them a day off. Even more so, PSAI insights could be used to bargain better for pay. Typically, in an asymmetrical power structure the workers have inferior bargaining power [464]. Thus, PSAI must be conceived to maximize the bargaining power each IW has. To empower an IW with such technology, we need to design beyond purposes of nudges and reflection [470] and design for collective bargaining [471]. Data–driven bargaining has its basis in traditional methods for Human Resource Management such as timekeeping [472]. As workers become more conscious of themselves due to increasing perceived or actual technological surveillance, they have the potential to contest claims by their employer. Ideally, such PSAI technologies must be accessible to IWs independent of their employer and independent of the PSAI their organization may have already deployed. If the outputs of PSAI technologies are only limited to actions IWs should do (e.g., "you seem very stressed, take a break"), they might not be able to accumulate enough knowledge of what they have been doing (e.g., "you have been overworked for 60 days, please consult your manager"). Future PSAI technologies need to make inferences that are reproducible and support sensemaking. Note, however, isolated individual understandings can fall be limited in an asymmetrical power structure [470, 471]. Instead, pooling of information also fit within the norms of distribution, as cumulatively sharing PSAI can help them gain better perspective on the algorithmic estimates. Research and activism on both crowdwork and gig–work have proposed to arm workers with data of their work to combat asymmetry [470, 473]. From our findings, the right iteration of PSAI could serve this purpose for IWs and help them build their own conceptions of these algorithmic estimates (section 8.2). Taking a leaf from studies in crowd–work [474], the next step would be to pursue research on collective platforms to leverage behavioral data for workplace bargaining.

9.4 Enabling Worker Consent

I wanted to emphasize that despite the directions for responsible practices from the technological perspective, organizations have full reign on establishing these technologies for nefarious purposes. A fundamental aspect of power asymmetry is *information asymmetry* [475] and PSAI is designed to produce new information. When this information is distributed inequitably among the stakeholders, it can exacerbate the asymmetry.

> "Unfortunately, all these approaches provide both a logical and a physical single point of control over our personal data: typically, they entail lodging information in the cloud where the service is running. This naturally *leads to a host of trust issues for users*, who find themselves not just having to trust the service directly but also the infrastructure providers involved"— Chaudhry *et al.* [476]

As researchers we are responsible for building the single points discussed above. One can argue that the user's choice is paramount, and therefore, the responsibility lies on the information worker. However, it is naive to assume that fair choice in an asymmetrical setting [477]. Even though workers provide a service, a capitalist society can position them as interchangeable and dependent on their employer. It is not unlikely that some workers can reject such systems (and by extension such organizations), but rarely does this transform into collective action [478]. Conversely, we researchers, and developers of these passive sensing frameworks, function independent of our research subjects' asymmetrical work paradigm. It puts us in a unique position to recommend change by encouraging new practices. Unfortunately, a recent study indicated that technologists are getting more and more decoupled from their responsibilities towards their data subjects or data providers [479]. A possible way out could be through the right set of guidelines to ensure that workers can consent in a more complete way. Chowdhary *et al.* proposes expanding the notion of informed consent for worker wellbeing technologies with *FRIES* — Freely-given, Reversible, Informed, Enthusiastic, and Specific [480]. Many of the recommendations to support this idea require workers to overcome technological barriers. In alignment with my own research, I believe that integrating technological features to support active reflection and assessment of organizations can not only protect workers' right to meaningful consent but also promise a more caring tool that helps increase their overall satisfaction.

9.5 Transferrability of Findings to Changing Landscapes of Work

All the studies I presented involved US–based information workers. To some extent the motivations of this research and even the interpretations of the findings are flavored with the socio-technical specifics of the US. However, the studies do vary in time. The first few studies in my dissertation began in 2018 (section 3.1) and the recruitment for the last study began in 2023 (section 8.2). If one were to juxtapose my studies with various changes in organizations' interest in thriving workers, one would find my studies reflective of these variations. Starting from a time when worker performance needed to be understood, to entering and exiting a pandemic which completely overhauled the way we worked. Some organizations prioritized worker wellbeing front-and-center and we saw these employers attract more information workers. When the writing of this dissertation was completed, many companies heavily reliant on information work heavy had come out of large-scale lay-offs. Many were still financially cornered. They were less interested in personalized wellbeing and cared more about meeting their productivity requirements. In this section I have speculated the implications of my research in work environments outside of the US and also, in different cycles of the economy.

9.5.1 Information Work Outside the USA

My dissertation was motivated by the importance of day–level improvements to support thriving information workers. The individual is a function of their environment and oftentimes the infrastructures around them can have a significant effect on their wellbeing. The labor policies and the underlying work–leisure culture of the US is unique. A good way to appreciate it is by contrasting it with the another equally developed economical context with a similar growth trajectory, Europe. Sometime in the late 1960's, we witnessed a decline in European working hours in terms of hours per week and the total weeks per year [481]. More recently, we are seeing Europe–based information workers gain more safeguards against detrimental wellbeing practices with "right to disconnnect" [482]. Several countries have adopted this idea with protective policies that can hold employers in violation if they communicate with their workers after a certain time [483]. In light of these, one might wonder what is utility of an information worker in Europe gaining additional insight into their wellbeing, especially through technologies like PSAI.

These European policies help create boundaries for workers to appreciate and actively utilize their leisure time. It is a luxury that only few US–based information workers have. Yet, these benefits come at a cost. The US is far ahead of Europe in terms of economic productivity [484]. Apart from losing competitive edge, this gap becomes concerning when labor might not be able to afford the wellbeing benefits it desires. Returning to the individual–level, surveys from 2022 show that workers in Europe are notably less engaged [485]. Therefore, the personalized insights produced by PSAI might still hold value in explaining to workers how and where they are likely to contribute the most while still maintaining their wellbeing needs.

Even if we can justify the technology, different work paradigms also lead to distinct socio-technical constraints. Much like the wellbeing–related policies, Europe has also established key policies to protect certain classes of employee data through *General Data Protection Regulation* (GDPR) [486]. Such regulatory policies can provide the necessary guardrails for designing PSAI. Not only can the data collection be limited but also the access to inferences and insights. Potentially, worker wellbeing tools might even become as personal as a one's health trackers. However, the regulation also introduces a paradox where it can become challenging to conduct the necessary research on PSAI in European contexts. Alternatively, other fast–developing economies with large information worker populations might not have policies equivalent to GDPR. It can be tempting to conduct research in these geo-politcal contexts but, if the findings from my US–based participants are indicative, the lack of all regulation will heighten worker concerns with PSAI.

Taken together, the motivations to use (and not use) PSAI and how the technology is appropriated is likely to differ based on different regulations and policy. Although the underlying technology might not drastically change, the way PSAI is instantiated will vary in terms of data source, flow, sharing, and stakeholders.

9.5.2 Information Work in Hard Times

The way an organization views its workers and the way workers view them back is often a function of the peripheral economy. Coincidentally, my dissertation spans two macroeconomic negative incidents that significantly changed the worker–organization relationship among information workers in the US. First, the COVID-19 pandemic induced safe–distancing requirements forced organizations to think about employee wellbeing. The idea of a successful thriving employee was centered around worker–needs [487]. However, sometime after the pandemic subsided, we witnessed an economic downturn with mass layoffs. Such organizational restructuring indicated a focus on enterprise–level productivity needs at the expense of individual worker performance or wellbeing [488]. Thus, in a short-span of time we retrospectively anticipate the shifting priorities in promoting technologies like PSAI depending on economy–wide circumstances.

In times when organizations are indifferent to workers' agency and personalized needs, they are trying to secure their short-term needs for economic survival. However, research shows that organizations that focus on workers' needs beyond financial incentive tend to recover better in the long run [489]. Technology like PSAI could play a role in highlighting their workers' state to help them take stock of workers again. However, the technology itself could also be misappropriated to disenfranchise the worker. When the job market is frugal with opportunity it is considered *austere*. Studies show that algorithmic profiling of job seekers in times of austerity is often considered desirable by the state but ends up being harmful for the individuals themselves [490]. When the market offers limited choices for the information worker to go, one can imagine organizations gain access to PSAI and identify workers that are predicted to perform highly without stressful bouts. Such motivations are only likely to increase disparities. A potential solution to this is for states to define statutory limits on decision making through PSAI in times of economic downturn.

Macroeconomic challenges are not the only temporal events that make information workers vulnerable to misuse of PSAI. Depending on a worker's experience, they might find themselves in roles with high *precariousness*, such as contractual work or temporary occupations. Unlike full-time roles, these employment contracts are far less secure. Prior literature has found extensive evidence that job precarity can worsen worker wellbeing [491]. It can be tempting to deploy PSAI to help workers understand their wellbeing better, but the anxieties of losing their job are likely to supersede any potential benefit of the technology. In fact, the introduction of such technology might be more consequential than other secure roles as their evaluations are directly tied to their contract. When their full-time counterparts discuss their insights from PSAI, they might just receive a different bonus, additional breaks, or a figurative "slap on the wrist". By contrast, their contract is more fragile and might be terminated. Even positive inferences can be interpreted by an organization as a reason to hire a worker with lower expectations. Therefore, the policies also need to regulate the rights of different workers when it comes to sharing data with PSAI and moreover involving their employer.

My studies on PSAI have discussed technological possibilities that sense complex behavioral dynamics to interpret insights of effectiveness. Essentially I demonstrated approaches to apply computational modeling and machine learning to reduce the human experience into consummable information. These insights should be considered starting points of dialogues between workers and their organizations. Much like other instruments of explaining human ability, PSAI also suffers from the challenge of being used for fallacious purposes like *reification* and *ranking* [492]. Organizations that fall into this trap will foster an economy of dissatisfied workers that will lead to long-term losses. Instead, I believe PSAI needs to be utilized as a tool for understanding, augmenting, and recovering the individual worker.

9.6 Who Monitors the Monitoring?

Most worker monitoring technologies are marketed for organizational — not personal — consumption [403, 404, 389]. The power asymmetry at work makes these information flows further opaque. One of the exceptions to this was *Viva Insights* [407] which at least allowed some joint–initiative — the organization might need to purchase or subscribe to the service but each individual IW gets the discretion to use the technology. Yet, these cases do not entirely alleviate IW's anxieties of privacy intrusion as we know from the use of health trackers in wellbeing incentive programs [493]. A recent multi-stakeholder analysis by Kawakami *et al.* showed that despite best intentions, even if developers and organizations were to follow better practices (section 9.3), the lack of accountability raises many important questions for the practical deployment of PSAI as empowering technology. [494]. Through this section, I bring to attention the role of stakeholders to provide checks and balances; (i) *Regulators* who can react to deployments of PSAI, and (ii) *Researchers* stakeholders who can preempt future PSAI.

9.6.1 Role of Regulators

The need for improved legislation of worker surveillance is not new [495], but the urgency at which it needs to be revised needs to match the rapid development (and deployment) of AI technologies [496]. Ajunwa *et al.*, have proposed an "Employee Privacy Protection Act" (EPPA) to limit data harnessed by technologies like PSAI at work [432]. They have also proposed an "Employee Health Information Privacy Act" (EHIPA) to tackle unscrupulous data transactions by third–parties, which could help mitigate some of the challenges to PSAI that senses phenomena exclusive of the workplace [432]. Such propositions are certainly a step in the right direction but are centered on confining the flow of the data, i.e., limiting whom they go to, but not how they use it. The algorithmic element of PSAI makes its mechanics elusive and therefore traditional auditing approaches will be lacking.

Assessing the impacts of PSAI despite the black box. Many PSAI technologies are shrouded as "black–boxes" and this opacity enables certain folk theories on the capabilities of these systems [391]. It is well known that explainability of machine–learning and AI systems is a hard problem, but I believe adding regulation can motivate developers of PSAI to account for the information flows along the dimensions of contextual norms in information work. We can follow the idea of *Model Cards* proposed by Mitchell *et al.* [497], to document intended usage of PSAI. For instance, developers might need to expand on how the algorithmic inferences could be consequential to an IW's employment with explicitly defined entry points for human–reappraisal and stakeholder involvement. Similarly, they could be required to disclose which aspects of wellbeing and performance are ignored by the system (e.g., "this PSAI cannot be used to infer your team management skills" or "this PSAI is not appropriate for communication–driven roles").

Assessing PSAI within socio-economic context. Grill and Andalibi had called to increase the visibility of the social impacts of algorithmic phenotyping [391]. Contemporary research has already raised the concerns surrounding the social dynamics of emotional recognition [467], a well–documented manifestations of PSAI. Ideally, regulations must protect against foreseeable, but anomalous, economic scenarios that compel organizational supervision. For instance, in the future, economic downturn can be used to justify the diversion of PSAI inferences for operational decisions such as downsizing. Organizations can argue these situations are analogous to PSAI for public–health [422]. These crisis scenarios require regulation the most. Auditors should be able to protect certain jobs that are considered more precarious from PSAI, e.g., contractual positions. At the same time, certain sectors might be deemed too austere for responsible utilization of PSAI. Sectors that lack sufficient alternative job openings lead to austerity that makes it illusory for workers to improve with PSAI or even meaningfully reject any enforced PSAI. Future research can illuminate other sociological factors that inform protective regulation. In this way, thinking of more worker-centric PSAI could ensure that individual liberties are upheld despite the necessary supervision required for organizational progress.

9.6.2 Role of Researchers

A harder question to answer is defining changes in the research of PSAI for workers. The path forward needs insiders to embrace reflexivity on our own methods, but also adopt calls for more human-centered approaches from "outsiders", who have critiqued this research. Oftentimes the scientific advancement of technology–supported HRM hinges on capturing and modeling otherwise unseen or ignored phenomena [111, 26, 498]. Sometimes this research is presented as morally indifferent to misuse. This indifference starts eroding when researchers start intersecting with more societal disciplines, such as HCI and CSCW. Yet, research projects and papers that do anticipate misuse are often limited to statements that urge for consented usage. Unfortunately, data subjects might not be able to take informed decisions without the appropriate disclosures of PSAI. Despite our worker–centric approach, in a technology–forward environment an IW's judgments could be clouded by their personal theories of AI as well as folk theories about the inner functioning of AI-based systems [391]. It is only when we appreciate external critique can we understand the risks of perpetuating PSAI, such as the potential for self–harm [499].

Participatory contributions to development of PSAI. The bare minimum would be to include reflective discussions based on the norms of information flow among information workers (or more specific norms suited to their subpopulation). A more worker–centric approach would be to embed qualitative methods such as the scenario–based interviews we conducted as a formative evaluation. Ideally, researchers should have IWs participate in the entire research life–cycle, drawing upon principles and ideas from participatory action research [500]. Even before IRB reviews, study protocols could be informed with feedback from IWs to understand if appropriate measures or phenomena are the input for PSAI. Later, models can be validated through a participatory lens where co–researcher IWs can vet the practical value or harms or potential harms. Studies like *WeBuildAI* [501] already provide some framework for participatory algorithmic decision making. Future work should expand this to algorithmic phenotyping.

External feedback for research on PSAI. Quantitative researchers also need to understand that participatory methods and qualitative evaluations will not create a universally accepted instance of PSAI. As Calacci notes that participatory algorithm design for workers might not be able to reconcile multiple conflicting stakeholders, but can at least ensure that normative expectations are not breached [470]. In practice, many researchers innovating new PSAI do not work on recruiting, data acquisition, or participant communication. After all, research on PSAI is often propelled by datasets of behavioral data because these are practical and desirable to support scientific replicability and reproducibility. However, these data also distance researchers from the data–subjects, and in some cases may lead to dehumanized conceptions of data–subjects and donors as simply "training data" or "numbers" [393]. To mitigate the impersonal relationship between researchers and data-subjects, we might consider setting up an independent advisory board formed of subjects and outsiders. Overall, increasing worker-centered research on PSAI can bridge this gap and produce more sensitive and humane systems to improve prosperity of IWs.

9.7 Conclusion

Passive sensing is an immensely powerful tool. As the amount of technology an urban worker interacts with increases, the opportunities to understand their behavior increases. Through this dissertation, I provided evidence that we can understand workers better by leveraging everyday digital technology to sense otherwise ignored phenomena related to work. Despite my optimism, I charted out the various challenges in making such applications meaningful for the workers they sense. The selling point of passive sensing is the promise of its "passive" nature. By design, these tools are unobtrusive, automatic, and continuous. However, these very value propositions can be weaponized against the humans they sense. My investigations have made me realize that the way forward may be a middle path. I imagine a future where Passive Sensing–enabled AI for workers is out of the way but can be glanced, studied, and even vetted. It is semi-automatic and encourages human input to define and label circumstances that cannot be measured. It can run as long as a worker wants but cannot operate indiscriminately. Eventually, the training wheels should come off and the worker can navigate work and life to achieve success.

REFERENCES

[1] S. Tangen, "Demystifying productivity and performance," *International Journal of Productivity and performance management*, 2005.

[2] C. C. Kuhlthau, "The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources," *Journal of the American Society for information Science*, vol. 50, no. 5, pp. 399–412, 1999.

[3] P. Warr and K. Nielsen, "Wellbeing and work performance," *Handbook of well-being. Salt Lake City, UT: DEF Publishers*, 2018.

[4] W. C. Borman, R. Bryant, and J. Dorio, "The measurement of task performance as criteria in selection research," *Handbook of employee selection*, pp. 439–461, 2010.

[5] N. C. Carpenter, C. M. Berry, and L. Houston, "A meta-analytic comparison of self-reported and other-reported organizational citizenship behavior," *Journal of Organizational Behavior*, vol. 35, no. 4, pp. 547–574, 2014.

[6] R. D. Goffin and I. R. Gellatly, "A multi-rater assessment of organizational commitment: Are self-report measures biased?" *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 22, no. 4, pp. 437–451, 2001.

[7] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, 1992.

[8] L. S. Aiken and S. G. West, "Invalidity of true experiments: Self-report pretest biases," *Evaluation review*, vol. 14, no. 4, pp. 374–390, 1990.

[9] I. Krumpal, "Determinants of social desirability bias in sensitive surveys: A literature review," *Quality & Quantity*, vol. 47, no. 4, pp. 2025–2047, 2013.

[10] G. Bagheri, R. Zarei, and M. N. Aeen, "Organizational silence (basic concepts and its development factors)," *Ideal Type of Management*, 2012.

[11] E. W. Morrison and F. J. Milliken, "Organizational silence: A barrier to change and development in a pluralistic world," *Academy of Management review*, 2000.

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

[13] V. P. Cornet and R. J. Holden, "Systematic review of smartphone-based passive sensing for health and wellbeing," *Journal of biomedical informatics*, vol. 77, pp. 120–132, 2018.

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

[15] V. Das Swain *et al.*, "Leveraging wifi network logs to infer student collocation and its relationship with academic performance," *EPJ Data Science*, vol. 12, no. 1, pp. 1–25, 2023.

[16] V. Das Swain, M. D. Reddy, K. A. Nies, L. Tay, M. De Choudhury, and G. D. Abowd, "Birds of a Feather Clock Together: A Study of Person–Organization Fit Through Latent Activity Routines," *Proc. ACM Hum.-Comput. Interact*, no. CSCW, 2019.

[17] V. Das Swain, K. Saha, M. D. Reddy, H. Rajvanshy, G. D. Abowd, and M. De Choudhury, "Modeling organizational culture with workplace experiences shared on glassdoor," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–15.

[18] V. Das Swain *et al.*, "Empirical networks for localized covid-19 interventions using wifi infrastructure at university campuses," *Frontiers in Digital Health*, vol. 5, p. 1 060 828, 2023.

[19] V. Das Swain, V. Chen, S. Mishra, S. M. Mattingly, G. D. Abowd, and M. De Choudhury, "Semantic gap in predicting mental wellbeing through passive sensing," in *Proceedings of the 2022 CHI conference on human factors in computing systems*, 2022.

[20] V. Das Swain, L. Gao, W. A. Wood, S. C. Matli, G. D. Abowd, and M. De Choudhury, "Algorithmic power or punishment: Information worker perspectives on passive sensing enabled ai phenotyping of performance and wellbeing," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–17.

[21] P. F. Drucker, *The effective executive*. Routledge, 2018.

[22] T. H. Davenport, *Thinking for a living: how to get better performances and results from knowledge workers*. Harvard Business Press, 2005.

[23] L. Yang *et al.*, "The effects of remote work on collaboration among information workers," *Nature human behaviour*, pp. 1–12, 2021.

[24] K. Saha, M. D. Reddy, S. Mattingly, E. Moskal, A. Sirigiri, and M. De Choudhury, "Libra: On linkedin based role ambiguity and its relationship with wellbeing and job performance," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–30, 2019.

[25] H. Cao *et al.*, "Large scale analysis of multitasking behavior during remote meetings," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–13.

[26] G. Mark, S. T. Iqbal, M. Czerwinski, and P. Johns, "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*, 2014, pp. 3025–3034.

[27] J. Cheng, J. Teevan, S. T. Iqbal, and M. S. Bernstein, "Break it down: A comparison of macro-and microtasks," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 4061–4064.

[28] S. Ricard, *The year of the knowledge worker, \NoCaseChange{https://www.forbes.com/sites/forbestechcouncil/2020/12/10/the-year-of-the-knowledge-worker/?sh=659e572f7fbb}*, Accessed: 2021-12-07, 2020.

[29] F. Gillet, *The global information worker population swells to 1.25 billion in 2018*, Accessed: 2021-12-07, 2019.

[30] P. F. Drucker, "Knowledge-worker productivity: The biggest challenge," *California management review*, vol. 41, no. 2, pp. 79–94, 1999.

[31] R. Erne, "What is productivity in knowledge work?-a cross-industrial view-.," *J. Univers. Comput. Sci.*, vol. 17, no. 10, pp. 1367–1389, 2011.

[32] M. Rotundo and P. R. Sackett, "The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach.," *Journal of applied psychology*, vol. 87, no. 1, p. 66, 2002.

[33] G. Anderson and C. Viswesvaran, "An update of the validity of personality scales in personnel selection: A meta-analysis of studies published after 1992," in *13th Annual Conference of the Society of Industrial and Organizational Psychology, Dallas*, 1998.

[34] M. R. Barrick and M. K. Mount, "The big five personality dimensions and job performance: A meta-analysis," *Personnel psychology*, vol. 44, no. 1, pp. 1–26, 1991.

[35] J. F. Salgado, "The five factor model of personality and job performance in the european community.," *Journal of Applied psychology*, vol. 82, no. 1, p. 30, 1997.

[36] R. P. Tett, D. N. Jackson, and M. Rothstein, "Personality measures as predictors of job performance: A meta-analytic review," *Personnel psychology*, vol. 44, no. 4, pp. 703–742, 1991.

[37] M. R. Barrick, M. K. Mount, and T. A. Judge, "Personality and performance at the beginning of the new millennium: What do we know and where do we go next?" *International Journal of Selection and assessment*, vol. 9, no. 1-2, pp. 9–30, 2001.

[38] D. Bartram, "The great eight competencies: A criterion-centric approach to validation.," *Journal of applied psychology*, vol. 90, no. 6, p. 1185, 2005.

[39] G. J. Feist, "A meta-analysis of personality in scientific and artistic creativity," *Personality and social psychology review*, vol. 2, no. 4, pp. 290–309, 1998.

[40] J. Colquitt, J. A. LePine, M. J. Wesson, and X. Wu, *Organizational behavior: Essentials for improving performance and commitment*. Mcgraw-Hill Irwin New York, NY, 2010.

[41] R. S. Dalal, D. P. Bhave, and J. Fiset, "Within-person variability in job performance: A theoretical review and research agenda," *Journal of Management*, vol. 40, no. 5, pp. 1396–1436, 2014.

[42] S. W. Kozlowski and B. M. Hults, "An exploration of climates for technical updating and performance," *Personnel psychology*, vol. 40, no. 3, pp. 539–563, 1987.

[43] G. Mark, V. M. Gonzalez, and J. Harris, "No task left behind? examining the nature of fragmented work," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005, pp. 321–330.

[44] G. M. Olson and J. S. Olson, "Distance matters," *Human–computer interaction*, vol. 15, no. 2-3, pp. 139–178, 2000.

[45] A. C. Edmondson, R. M. Bohmer, and G. P. Pisano, "Disrupted routines: Team learning and new technology implementation in hospitals," *Administrative science quarterly*, vol. 46, no. 4, pp. 685–716, 2001.

[46] E. H. Trainer, A. Kalyanasundaram, C. Chaihirunkarn, and J. D. Herbsleb, "How to hackathon: Socio-technical tradeoffs in brief, intensive collocation," in *proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, 2016, pp. 1118–1130.

[47] R. T. Sparrowe, R. C. Liden, S. J. Wayne, and M. L. Kraimer, "Social networks and the performance of individuals and groups," *Academy of management journal*, vol. 44, no. 2, pp. 316–325, 2001.

[48] G. R. Jones, "Psychological orientation and the process of organizational socialization: An interactionist perspective," *Academy of management review*, vol. 8, no. 3, pp. 464–474, 1983.

[49] P. M. Muchinsky and C. J. Monahan, "What is person-environment congruence? supplementary versus complementary models of fit," *Journal of vocational behavior*, vol. 31, no. 3, pp. 268–277, 1987.

[50] C. A. O'Reilly III, J. Chatman, and D. F. Caldwell, "People and organizational culture: A profile comparison approach to assessing person-organization fit," *Academy of management journal*, vol. 34, no. 3, pp. 487–516, 1991.

[51] R. Vaitenas and Y. Wiener, "Developmental, emotional, and interest factors in voluntary mid-career change," *Journal of Vocational Behavior*, vol. 11, no. 3, pp. 291–304, 1977.

[52] J. E. Dittrich and M. R. Carrell, "Organizational equity perceptions, employee job satisfaction, and departmental absence and turnover rates," *Organizational behavior and human performance*, vol. 24, no. 1, pp. 29–40, 1979.

[53] T. P. Summers and W. H. Hendrix, "Modelling the role of pay equity perceptions: A field study," *Journal of Occupational Psychology*, vol. 64, no. 2, pp. 145–157, 1991.

[54] D. F. Caldwell and C. A. O'Reilly III, "Measuring person-job fit with a profile-comparison process.," *Journal of applied psychology*, vol. 75, no. 6, p. 648, 1990.

[55] D. Dawda and J. Martin, "Psychologists on psychology: The inquiry beliefs of psychologists by subdiscipline, gender, and age," *Review of General Psychology*, vol. 5, no. 3, pp. 163–179, 2001.

[56] J. R. Edwards, D. M. Cable, I. O. Williamson, L. S. Lambert, and A. J. Shipp, "The phenomenology of fit: Linking the person and environment to the subjective experience of person-environment fit.," *Journal of Applied Psychology*, vol. 91, no. 4, p. 802, 2006.

[57] B. M. Berger, *An essay on culture: Symbolic structure and social structure*. Univ of California Press, 1995.

[58] J. Silvester, N. R. Anderson, and F. Patterson, "Organizational culture change: An inter-group attributional analysis," *J. Occup. Organ. Psychol*, 1999.

[59] R. A. Cooke and D. M. Rousseau, "Behavioral norms and expectations: A quantitative approach to the assessment of organizational culture," *Group & Organization Studies*, vol. 13, no. 3, pp. 245–273, 1988.

[60] R. Strack, C. Von Der Linden, M. Booker, and A. Strohmayr, "Decoding global talent," *BCG Perspectives*, 2014.

[61] G. Hofstede, *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications, 2001.

[62] C. Yilmaz and E. Ergun, "Organizational culture and firm effectiveness: An examination of relative effects of culture traits and the balanced culture hypothesis in an emerging economy," *Journal of world business*, vol. 43, no. 3, pp. 290–306, 2008.

[63] Y. Vardi, "The effects of organizational and ethical climates on misconduct at work," *Journal of Business ethics*, vol. 29, no. 4, pp. 325–337, 2001.

[64] G. Hofstede, B. Neuijen, D. D. Ohayv, and G. Sanders, "Measuring organizational cultures: A qualitative and quantitative study across twenty cases," *Administrative science quarterly*, 1990.

[65] D. I. Prajogo and C. M. McDermott, "The relationship between multidimensional organizational culture and performance," *Int. J. Oper. Prod. Manag.*, 2011.

[66] Y. Baruch and B. C. Holtom, "Survey response rate levels and trends in organizational research," *Human relations*, vol. 61, no. 8, pp. 1139–1160, 2008.

[67] W. H. Organization *et al.*, "Critical preparedness, readiness and response actions for covid-19: Interim guidance, 22 march 2020," World Health Organization, Tech. Rep., 2020.

[68] C. for Disease Control and Prevention, *Community npis: Flu prevention in community settings*, Available at: https://www.cdc.gov/nonpharmaceutical-interventions/community/index.html, 2020.

[69] S. Friedman, T. Hurley, T. Fishman, and P. Fritz, *Covid-19 impact on higher education*, Available at: https://www2.deloitte.com/us/en/pages/public-sector/articles/covid-19-impact-on-higher-education.html, 2020.

[70] A. DePietro, *Here's a look at the impact of coronavirus (covid-19) on colleges and universities in the u.s.* Available at: https://www.forbes.com/sites/andrewdepietro/2020/04/30/impact-coronavirus-covid-19-colleges-universities/?sh=26c3f79c61a6, 2020.

[71] E. Dorn, B. Hancock, J. Sarakatsannis, and E. Viruleg, "Covid-19 and student learning in the united states: The hurt could last a lifetime," *McKinsey & Company*, 2020.

[72] J. P. Azevedo, A. Hasan, D. Goldemberg, S. A. Iqbal, and K. Geven, *Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: A set of global estimates*. The World Bank, 2020.

[73] I. Chirikov, K. M. Soria, B. Horgos, and D. Jones-White, "Undergraduate and graduate students' mental health during the covid-19 pandemic," 2020.

[74] C. Woolston, *Signs of depression and anxiety soar among us graduate students during pandemic*, Available at: https://www.nature.com/articles/d41586-020-02439-6, 2020.

[75] T. Choudhury *et al.*, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.

[76] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Medical engineering & physics*, vol. 30, no. 10, pp. 1364–1386, 2008.

[77] H. Lu *et al.*, "Stresssense: Detecting stress in unconstrained acoustic environments using smartphones," in *Proceedings of the 2012 ACM conference on ubiquitous computing*, 2012, pp. 351–360.

[78] R. Wang *et al.*, "Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Ubicomp*, ACM, 2014, pp. 3–14.

[79] K. Saha, L. Chan, K. De Barbaro, G. D. Abowd, and M. De Choudhury, "Inferring mood instability on social media by leveraging ecological momentary assessments," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 95, 2017.

[80] G. Mark, S. T. Iqbal, M. Czerwinski, P. Johns, A. Sano, and Y. Lutchyn, "Email duration, batching and self-interruption: Patterns of email use on productivity and stress," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 1717–1728.

[81] F. Schaule, J. O. Johanssen, B. Bruegge, and V. Loftness, "Employing consumer wearables to detect office workers' cognitive load for interruption management," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–20, 2018.

[82] S. Mirjafari *et al.*, "Differentiating higher and lower job performers in the workplace using mobile sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, p. 37, 2019.

[83] H. Kaur, A. C. Williams, D. McDuff, M. Czerwinski, J. Teevan, and S. T. Iqbal, "Optimizing for happiness and productivity: Modeling opportune moments for transitions and breaks at work," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–15.

[84] S. Nepal *et al.*, "Assessing the impact of commuting on workplace performance using mobile sensing," *IEEE Pervasive Computing*, vol. 20, no. 4, pp. 52–60, 2021.

[85] P. Lukowicz, S. Pentland, and A. Ferscha, "From context awareness to socially aware computing," *IEEE pervasive computing*, vol. 11, no. 1, pp. 32–41, 2011.

[86] D. O. Olguín and A. S. Pentland, "Social sensors for automatic data collection," *AMCIS 2008 Proceedings*, p. 171, 2008.

[87] D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 43–55, 2009.

[88] D. Olguín Olguín, "Sensor-based organizational design and engineering," Ph.D. dissertation, Massachusetts Institute of Technology, 2011.

[89] C. Brown *et al.*, "The architecture of innovation: Tracking face-to-face interactions with ubicomp technologies," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2014, pp. 811–822.

[90] A. Montanari, C. Mascolo, K. Sailer, and S. Nawaz, "Detecting emerging activity-based working traits through wearable technology," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 86, 2017.

[91] V. Prabhakaran, H. Neralwala, O. Rambow, and M. T. Diab, "Annotations for power relations on email threads.," in *LREC*, 2012.

[92] J. B. Begole, J. C. Tang, and R. Hill, "Rhythm modeling, visualizations and applications," in *Proceedings of the 16th annual ACM symposium on User interface software and technology*, ACM, 2003, pp. 11–20.

[93] N. Eagle and A. S. Pentland, "Eigenbehaviors: Identifying structure in routine," *Behavioral Ecology and Sociobiology*, vol. 63, no. 7, pp. 1057–1066, 2009.

[94] K. Saha *et al.*, “Social media as a passive sensor in longitudinal studies of human behavior and wellbeing,” in *CHI Ext. Abstracts*, ACM, 2019.

[95] A. Archambault and J. Grudin, “A longitudinal study of facebook, linkedin, & twitter use,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 2741–2750.

[96] H. Zhang, M. De Choudhury, and J. Grudin, “Creepy but inevitable?: The evolution of social networking,” in *Proc. CSCW*, ACM, 2014.

[97] J. DiMicco, D. R. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller, “Motivations for social networking at work,” in *CSCW*, 2008.

[98] W. Geyer, C. Dugan, J. DiMicco, D. R. Millen, B. Brownholtz, and M. Muller, “Use and reuse of shared lists as a social content type,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2008, pp. 1545–1554.

[99] N. S. Shami *et al.*, “Understanding employee social media chatter with enterprise social pulse,” in *Proc. CSCW*, 2014.

[100] A. Golestani, M. Masli, N. S. Shami, J. Jones, A. Menon, and J. Mondal, “Real-time prediction of employee engagement using social media and text mining,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, 2018, pp. 1383–1387.

[101] N. S. Shami, M. Muller, A. Pal, M. Masli, and W. Geyer, “Inferring employee engagement from social media,” in *Proc. CHI*, 2015.

[102] S. Ghoshray, “Employer surveillance versus employee privacy: The new reality of social media and workplace privacy,” *N. Ky. L. Rev.*, 2013.

[103] W. S. Jacobson and S. H. Tufts, “To post or not to post: Employee rights and social media,” *Review of public personnel administration*, 2013.

[104] M. M. Skeels and J. Grudin, “When social networks cross boundaries: A case study of workplace use of facebook and linkedin,” in *GROUP*, 2009.

[105] T. Choudhury and A. Pentland, “Characterizing social networks using the sociometer,” *Proceedings of the North American association of computational social and organizational science (NAACSOS)*, 2004.

[106] N. Eagle and A. S. Pentland, “Reality mining: Sensing complex social systems,” *Personal and ubiquitous computing*, vol. 10, no. 4, pp. 255–268, 2006.

[107] P. Sapiezynski, A. Stopczynski, R. Gatej, and S. Lehmann, "Tracking human mobility using wifi signals," *PloS one*, vol. 10, no. 7, e0130824, 2015.

[108] M. H. S. Eldaw, M. Levene, and G. Roussos, "Presence analytics: Making sense of human social presence within a learning environment," in *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, IEEE, 2018, pp. 174–183.

[109] S. Ware *et al.*, "Large-scale automatic depression screening using meta-data from wifi infrastructure," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–27, 2018.

[110] A. Trivedi, C. Zakaria, R. Balan, A. Becker, G. Corey, and P. Shenoy, "Wifitrace: Network-based contact tracing for infectious diseases using passive wifi sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 1, pp. 1–26, 2021.

[111] S. M. Mattingly *et al.*, "The tesserae project: Large-scale, longitudinal, in situ, multimodal sensing of information workers," in *CHI Ext. Abstracts*, 2019.

[112] G. Mark, S. Iqbal, M. Czerwinski, and P. Johns, "Capturing the mood: Facebook and face-to-face encounters in the workplace," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, 2014, pp. 1082–1094.

[113] R. Purta *et al.*, "Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits," in *Proceedings of the 2016 ACM international symposium on wearable computers*, ACM, 2016, pp. 28–35.

[114] C. Viswesvaran and D. S. Ones, "Perspectives on models of job performance," *International Journal of Selection and Assessment*, vol. 8, no. 4, pp. 216–226, 2000.

[115] W. C. Borman and S. Motowidlo, "Expanding the criterion domain to include elements of contextual performance," *Personnel Selection in Organizations; San Francisco: Jossey-Bass*, p. 71, 1993.

[116] M. A. Griffin, A. Neal, and S. K. Parker, "A new model of work role performance: Positive behavior in uncertain and interdependent contexts," *Academy of management journal*, vol. 50, no. 2, pp. 327–347, 2007.

[117] L. J. Williams and S. E. Anderson, "Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors," *Journal of management*, vol. 17, no. 3, pp. 601–617, 1991.

[118] R. J. Bies, *Organizational citizenship behavior: The good soldier syndrome*, 1989.

[119] S. J. Motowidlo, "Job performance," *Handbook of psychology: Industrial and organizational psychology*, vol. 12, pp. 39–53, 2003.

[120] D. W. Organ, "Organizational citizenship behavior: It's construct clean-up time," *Human performance*, vol. 10, no. 2, pp. 85–97, 1997.

[121] N. Schmitt, J. M. Cortina, M. J. Ingerick, and D. Wiechmann, "Personnel selection and employee performance," *Handbook of psychology: Industrial and organizational psychology*, vol. 12, pp. 77–105, 2003.

[122] P. M. Podsakoff, S. B. MacKenzie, R. H. Moorman, and R. Fetter, "Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors," *The leadership quarterly*, vol. 1, no. 2, pp. 107–142, 1990.

[123] R. H. Moorman, "The influence of cognitive and affective based job satisfaction measures on the relationship between satisfaction and organizational citizenship behavior," *Human relations*, vol. 46, no. 6, pp. 759–776, 1993.

[124] S. Fox, P. E. Spector, A. Goh, K. Bruursema, and S. R. Kessler, "The deviant citizen: Measuring potential positive relations between counterproductive work behaviour and organizational citizenship behaviour," *Journal of Occupational and Organizational Psychology*, vol. 85, no. 1, pp. 199–220, 2012.

[125] R. S. Dalal, H. Lam, H. M. Weiss, E. R. Welch, and C. L. Hulin, "A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counterproductivity associations, and dynamic relationships with affect and overall job performance," *Academy of Management Journal*, vol. 52, no. 5, pp. 1051–1066, 2009.

[126] J. P. Campbell and B. M. Wiernik, "The modeling and assessment of work performance," *Annu. Rev. Organ. Psychol. Organ. Behav.*, vol. 2, no. 1, pp. 47–74, 2015.

[127] P. R. Sackett, "The structure of counterproductive work behaviors: Dimensionality and relationships with facets of job performance," *International journal of selection and assessment*, vol. 10, no. 1-2, pp. 5–11, 2002.

[128] R. J. Bennett and S. L. Robinson, "Development of a measure of workplace deviance.," *Journal of applied psychology*, vol. 85, no. 3, p. 349, 2000.

[129] K. Danna and R. W. Griffin, "Health and well-being in the workplace: A review and synthesis of the literature," *Journal of management*, vol. 25, no. 3, pp. 357–384, 1999.

[130] H. M. Davey, A. L. Barratt, P. N. Butow, and J. J. Deeks, "A one-item question with a likert or visual analog scale adequately measured current anxiety," *Journal of clinical epidemiology*, vol. 60, no. 4, pp. 356–360, 2007.

[131] R. S. Lazarus, "Psychological stress and the coping process.," 1966.

[132] S. Sonnentag and M. Frese, "Stress in organizations," *Comprehensive handbook of psychology*, vol. 12, pp. 453–491, 2003.

[133] Garmin, *Garmin health api*, Accessed: 2018-11-01, 2018.

[134] F. T. Stress and R. W. Paper, *NoCaseChange*{*https://assets.firstbeat.com/firstbeat/uploads/2015/11/Stress-and-recovery_white-paper_20145.pdf*}, Accessed: 2019-04-01, 2014.

[135] F. Technologies, "Automated fitness level (vo2max) estimation with heart rate and speed data," *https://hubloq.net/cloud/runlog/diverse/white_paper_VO2max_11-11-20142.pdf*, 2014.

[136] G. L. Kraft and R. A. Roberts, "Validation of the garmin forerunner 920 xt fitness watch vo 2 peak test," *Int. J. Innov. Educ. Res*, vol. 5, pp. 62–67, 2017.

[137] K. Saha, M. D. Reddy, S. M. Mattingly, E. Moskal, A. Sirigiri, and M. De Choudhury, "LibRA : On LinkedIn based Role Ambiguity and Its Relationship with Wellbeing and Job Performance," *Proc. ACM Hum.-Comput. Interact.*, no. CSCW, 2019.

[138] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. Al'Absi, and S. Shah, "Autosense: Unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, 2011, pp. 274–287.

[139] Google, "Activity recognition api," *https://developers.google.com/location-context/activity-recognition/*, 2018, Accessed: 2018-11-01.

[140] Gimbal, "Manager rest api," *https://docs.gimbal.com/rest.html*, 2018, Accessed: 2018-11-01.

[141] M. Bin Morshed *et al.*, "Prediction of mood instability with passive sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, 2019.

[142] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language use: Our words, our selves," *Annual review of psychology*, vol. 54, no. 1, pp. 547–577, 2003.

[143] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[144] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55–60.

[145] G. Coppersmith, C. Harman, and M. Dredze, "Measuring post traumatic stress disorder in twitter," in *Eighth international AAAI conference on weblogs and social media*, 2014.

[146] K. Saha *et al.*, "Imputing missing social media data stream in multisensor studies of human behavior," *linguistic analysis*, vol. 19, no. 29, p. 43,

[147] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing twitter" big data" for automatic emotion identification," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, IEEE, 2012, pp. 587–592.

[148] S. Aman and S. Szpakowicz, "Using roget's thesaurus for fine-grained emotion recognition," in *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*, 2008.

[149] C. Zakaria, A. Trivedi, E. Cecchet, M. Chee, P. Shenoy, and R. Balan, "Analyzing the impact of covid-19 control policies on campus occupancy and mobility via wifi sensing," *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, vol. 8, no. 3, pp. 1–26, 2022.

[150] Cisco, "Cisco wi-fi location-based services 4.1 design guide white paper," *https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Mobility/WiFiLBS-DG/wifich2.html*, 2014, Accessed: 2020-05-10.

[151] Accuware, *Accuware wi-fi location monitor*, Accessed: 2020-05-10, 2016.

[152] M. B. Kjaergaard and P. Nurmi, "Challenges for social sensing using wifi signals," in *Proceedings of the 1st ACM workshop on Mobile systems for computational social science*, 2012, pp. 17–21.

[153] W. Dunn, C. Brown, and A. McGuigan, "The ecology of human performance: A framework for considering the effect of context," *American Journal of Occupational Therapy*, vol. 48, no. 7, pp. 595–607, 1994.

[154] C. Sansone, C. C. Morf, and A. T. Panter, *The Sage handbook of methods in social psychology*. Sage Publications, 2003.

[155] I. Ajzen, "Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology," in *Advances in experimental social psychology*, vol. 20, Elsevier, 1987, pp. 1–63.

[156] D. S. Ones, "Personality at work: Raising awareness and correcting misconceptions," *Human Performance*, vol. 18, no. 4, pp. 389–404, 2005.

[157] L. M. Hough and F. L. Oswald, "Personality testing and industrial–organizational psychology: Reflections, progress, and prospects," *Industrial and Organizational Psychology*, vol. 1, no. 3, pp. 272–290, 2008.

[158] B. W. Roberts and D. Mroczek, "Personality trait change in adulthood," *Current directions in psychological science*, vol. 17, no. 1, pp. 31–35, 2008.

[159] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

[160] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on knowledge and data engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[161] J. Neter, *Applied linear statistical models*, vol. 4.

[162] S. E. Woo, A. T. Jebb, L. Tay, and S. Parrigon, "Putting the "person" in the center: Review and synthesis of person-centered approaches and methods in organizational science," *Organizational Research Methods*, vol. 21, no. 4, pp. 814–845, 2018.

[163] M. Wang and P. J. Hanges, "Latent class procedures: Applications to organizational research," *Organizational Research Methods*, vol. 14, no. 1, pp. 24–31, 2011.

[164] R. J. Foti, N. J. Thompson, and S. F. Allgood, "The pattern-oriented approach: A framework for the experience of work," *Industrial and Organizational Psychology*, vol. 4, no. 1, pp. 122–125, 2011.

[165] A. S. Gabriel, M. A. Daniels, J. M. Diefendorff, and G. J. Greguras, "Emotional labor actors: A latent profile analysis of emotional labor strategies.," *Journal of Applied Psychology*, vol. 100, no. 3, p. 863, 2015.

[166] E. Kreyszig, *Advanced engineering mathematics*. John Wiley & Sons, 2010.

[167] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[168] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[169] R. Bellman, *Dynamic programming*. Courier Corporation, 2013.

[170] R. P. Burns and R. Burns, *Business research methods and statistics using SPSS*. Sage, 2008.

[171] P. T. Costa Jr, J. H. Herbst, R. R. McCrae, J. Samuels, and D. J. Ozer, "The replicability and utility of three personality types," *European Journal of Personality*, vol. 16, no. S1, S73–S87, 2002.

[172] M. Gerlach, B. Farb, W. Revelle, and L. A. N. Amaral, "A robust data-driven approach identifies four personality types across four large data sets," *Nature Human Behaviour*, vol. 2, no. 10, p. 735, 2018.

[173] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2011, pp. 143–146.

[174] M. Kay and J. Wobbrock, "Artool," *https://cran.r-project.org/web/packages/ARTool/ARTool.pdf*, 2019, Accessed: 2019-04-27.

[175] D. L. Ferris, D. J. Brown, and D. Heller, "Organizational supports and organizational deviance: The mediating role of organization-based self-esteem," *Organizational Behavior and Human Decision Processes*, vol. 108, no. 2, pp. 279–286, 2009.

[176] J. Cohen, "Eta-squared and partial eta-squared in fixed factor anova designs," *Educational and psychological measurement*, vol. 33, no. 1, pp. 107–112, 1973.

[177] T. R. Levine and C. R. Hullett, "Eta squared, partial eta squared, and misreporting of effect size in communication research," *Human Communication Research*, vol. 28, no. 4, pp. 612–625, 2002.

[178] D. Lüdecke, "Sj stats," *https://cran.r-project.org/web/packages/sjstats/sjstats.pdf*, 2019, Accessed: 2019-04-27.

[179] J. L. Price and C. W. Mueller, *A Casual Model of Turnover Estimated for Nurses*. 1979.

[180] A. N. Kluger, "Commute variability and strain," *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 19, no. 2, pp. 147–165, 1998.

[181] R. W. Novaco, W. Kliewer, and A. Broquet, "Home environmental consequences of commute travel impedance," *American Journal of Community Psychology*, vol. 19, no. 6, pp. 881–909, 1991.

[182] Q. R. Jett and J. M. George, "Work interrupted: A closer look at the role of interruptions in organizational life," *Academy of management Review*, vol. 28, no. 3, pp. 494–507, 2003.

[183] A. Baethge and T. Rigotti, "Interruptions to workflow: Their relationship with irritation and satisfaction with performance, and the mediating roles of time pressure and mental demands," *Work & Stress*, vol. 27, no. 1, pp. 43–63, 2013.

[184] J. G. Bernburg and T. Thorlindsson, "Routine activities in social context: A closer look at the role of opportunity in deviant behavior," *Justice Quarterly*, vol. 18, no. 3, pp. 543–567, 2001.

[185] M. A. Konovsky and S. D. Pugh, "Citizenship behavior and social exchange," *Academy of management journal*, vol. 37, no. 3, pp. 656–669, 1994.

[186] L. L. Dahlberg and E. G. Krug, "Violence: A global public health problem," *Ciência & Saúde Coletiva*, vol. 11, pp. 1163–1178, 2006.

[187] D. O. Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 43–55, 2008.

[188] R. Wang, G. Harari, P. Hao, X. Zhou, and A. T. Campbell, "Smartgpa: How smartphones can assess and predict academic performance of college students," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 295–306.

[189] S. Das, S. Chatterjee, S. Chakraborty, and B. Mitra, "Groupsense: A lightweight framework for group identification," *IEEE Transactions on Mobile Computing*, vol. 18, no. 12, pp. 2856–2870, 2018.

[190] T. Nguyen, D. Phung, S. Gupta, and S. Venkatesh, "Extraction of latent patterns and contexts from social honest signals using hierarchical dirichlet processes," in

2013 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, 2013, pp. 47–55.

[191] H. Hong, C. Luo, and M. C. Chan, "Socialprobe: Understanding social interaction through passive wifi monitoring," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ACM, 2016, pp. 94–103.

[192] J. Shi, L. Meng, A. Striegel, C. Qiao, D. Koutsonikolas, and G. Challen, "A walk on the client side: Monitoring enterprise wifi networks using smartphone channel scans," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, IEEE, 2016, pp. 1–9.

[193] S. Geister, U. Konradt, and G. Hertel, "Effects of process feedback on motivation, satisfaction, and performance in virtual teams," *Small group research*, vol. 37, no. 5, pp. 459–489, 2006.

[194] R. Fruchter, P. Bosch-Sijtsema, and V. Ruohomäki, "Tension between perceived collocation and actual geographic distribution in project teams," *Ai & Society*, vol. 25, no. 2, pp. 183–192, 2010.

[195] A. Edmondson, "Psychological safety and learning behavior in work teams," *Administrative science quarterly*, vol. 44, no. 2, pp. 350–383, 1999.

[196] K. A. Jehn and E. A. Mannix, "The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance," *Academy of management journal*, vol. 44, no. 2, pp. 238–251, 2001.

[197] G. S. Van der Vegt, B. J. Emans, and E. Van De Vliert, "Patterns of interdependence in work teams: A two-level investigation of the relations with job and team satisfaction," *Personnel Psychology*, vol. 54, no. 1, pp. 51–69, 2001.

[198] M. L. Loughry, M. W. Ohland, and D. DeWayne Moore, "Development of a theory-based assessment of team member effectiveness," *Educational and psychological measurement*, vol. 67, no. 3, pp. 505–524, 2007.

[199] P. J. Carnevale and T. M. Probst, "Social values and social conflict in creative problem solving and categorization.," *Journal of personality and social psychology*, vol. 74, no. 5, p. 1300, 1998.

[200] S. E. Taylor and J. D. Brown, "Illusion and well-being: A social psychological perspective on mental health.," *Psychological bulletin*, vol. 103, no. 2, p. 193, 1988.

[201] J. Y. Jiang, X. Zhang, and D. Tjosvold, "Emotion regulation as a boundary condition of the relationship between team conflict and performance: A multi-level

examination," *Journal of Organizational Behavior*, vol. 34, no. 5, pp. 714–734, 2013.

[202] K. A. Jehn, "A qualitative analysis of conflict types and dimensions in organizational groups," *Administrative science quarterly*, pp. 530–557, 1997.

[203] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz, "A regularity statistic for medical data analysis," *Journal of clinical monitoring*, vol. 7, no. 4, pp. 335–345, 1991.

[204] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 329.

[205] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[206] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[207] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, p. 066 138, 2004.

[208] T. Chai and R. R. Draxler, "Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature," *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[209] M. Nefzger and J. Drasgow, "The needless assumption of normality in pearson's r.," *American Psychologist*, vol. 12, no. 10, p. 623, 1957.

[210] G. Y. Zou, "Toward using confidence intervals to compare correlations.," *Psychological methods*, vol. 12, no. 4, p. 399, 2007.

[211] J. A. Chatman, "Matching people and organizations: Selection and socialization in public accounting firms.," in *Academy of Management proceedings*, Academy of Management Briarcliff Manor, NY 10510, vol. 1989, 1991, pp. 199–203.

[212] D. M. Cable and T. A. Judge, "Person–organization fit, job choice decisions, and organizational entry," *Organizational behavior and human decision processes*, vol. 67, no. 3, pp. 294–311, 1996.

[213] D. M. Cable and J. R. Edwards, "Complementary and supplementary fit: A theoretical and empirical integration.," *Journal of applied psychology*, vol. 89, no. 5, p. 822, 2004.

[214] J. R. Edwards and M. E. Parry, "On the use of spline regression in the study of congruence in organizational research," *Organizational Research Methods*, vol. 21, no. 1, pp. 68–110, 2018.

[215] S. Liu, Y. Jiang, and A. Striegel, "Face-to-face proximity estimationusing bluetooth on smartphones," *IEEE Transactions on Mobile Computing*, vol. 13, no. 4, pp. 811–823, 2014.

[216] A. Stopczynski, J. E. Larsen, S. Lehmann, L. Dynowski, and M. Fuentes, "Participatory bluetooth sensing: A method for acquiring spatio-temporal data about participant mobility and interactions at large scale events," in *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, IEEE, 2013, pp. 242–247.

[217] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos, "Getting closer: An empirical investigation of the proximity of user to their smart phones," in *Proceedings of the 13th international conference on Ubiquitous computing*, ACM, 2011, pp. 163–172.

[218] A. L. Kristof-Brown, R. D. Zimmerman, and E. C. Johnson, "Consequences of individuals'fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit," *Personnel psychology*, vol. 58, no. 2, pp. 281–342, 2005.

[219] M. Turk, "A random walk through eigenspace," *IEICE transactions on information and systems*, vol. 84, no. 12, pp. 1586–1595, 2001.

[220] A. A. Miranda, Y.-A. Le Borgne, and G. Bontempi, "New routes from minimal approximation error to principal components," *Neural Processing Letters*, vol. 27, no. 3, pp. 197–207, 2008.

[221] R. M. Dawes and B. Corrigan, "Linear models in decision making.," *Psychological bulletin*, vol. 81, no. 2, p. 95, 1974.

[222] C. J. Soto and O. P. John, "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*, vol. 113, no. 1, p. 117, 2017.

[223] B. Schneider, "The people make the place," *Personnel psychology*, vol. 40, no. 3, pp. 437–453, 1987.

[224] M. S. Feldman and B. T. Pentland, "Reconceptualizing organizational routines as a source of flexibility and change," *Administrative science quarterly*, vol. 48, no. 1, pp. 94–118, 2003.

[225] D. A. Harrison, S. Mohammed, J. E. McGrath, A. T. Florey, and S. W. Vanderstoep, "Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality," *Personnel Psychology*, vol. 56, no. 3, pp. 633–669, 2003.

[226] L. Pérez-Nordtvedt, G. T. Payne, J. C. Short, and B. L. Kedia, "An entrainment-based model of temporal organizational fit, misfit, and performance," *Organization Science*, vol. 19, no. 5, pp. 785–801, 2008.

[227] J. Lorber, "Good patients and problem patients: Conformity and deviance in a general hospital," *Journal of health and social behavior*, pp. 213–225, 1975.

[228] S. Arbour, C. T. Kwantes, J. M. Kraft, and C. A. Boglarsky, "Person-organization fit: Using normative behaviors to predict workplace satisfaction, stress and intentions to stay," *Journal of Organizational Culture, Communications and Conflict*, vol. 18, no. 1, p. 41, 2014.

[229] M. Siegall and T. McDonald, "Person-organization value congruence, burnout and diversion of resources," *Personnel Review*, vol. 33, no. 3, pp. 291–301, 2004.

[230] U. Schallberger, "The influence of personality-characteristics on self-reports of working-conditions," *Zeitschrift fur Experimentelle Psychologie*, vol. 42, no. 1, pp. 111–131, 1995.

[231] J. G. Grzywacz, D. S. Carlson, and S. Shulkin, "Schedule flexibility and stress: Linking formal flexible arrangements and perceived flexibility to employee health," *Community, Work and Family*, vol. 11, no. 2, pp. 199–214, 2008.

[232] E. Jeffrey Hill, J. I. Jacob, L. L. Shannon, R. T. Brennan, V. L. Blanchard, and G. Martinengo, "Exploring the relationship of workplace flexibility, gender, and life stage to family-to-work conflict, and stress and burnout," *Community, Work and Family*, vol. 11, no. 2, pp. 165–181, 2008.

[233] D. F. Halpern, "How time-flexible work policies can reduce stress, improve health, and save money," *Stress and health*, vol. 21, no. 3, pp. 157–168, 2005.

[234] A. Grönlund, "Employee control in the era of flexibility: A stress buffer or a stress amplifier?" *European Societies*, vol. 9, no. 3, pp. 409–428, 2007.

[235] S. R. Barley, G. W. Meyer, and D. C. Gash, "Cultures of culture: Academics, practitioners and the pragmatics of normative control," *Administrative science quarterly*, pp. 24–60, 1988.

[236] A. Chamberlain, "Does company culture pay off? analyzing stock performance of'best places to work'companies," Glassdoor Research Report, March 2015, Tech. Rep., 2015.

[237] C. O'Reilly, "Corporations, culture, and commitment: Motivation and social control in organizations," *California management review*, 1989.

[238] R. A. Cooke and J. L. Szumal, "Using the organizational culture inventory to understand the operating cultures of organizations," *Handbook of organizational culture and climate*, 2000.

[239] S. R. Glaser, S. Zamanou, and K. Hacker, "Measuring and interpreting organizational culture," *Management communication quarterly*, 1987.

[240] R. E. Quinn and J. Rohrbaugh, "A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis," *Management science*, vol. 29, no. 3, pp. 363–377, 1983.

[241] Glassdoor, "Glassdoor give to get policy," *help.glassdoor.com/article/Give-to-get-policy/en_US*, 2018, Accessed: 2019-04-04.

[242] G. C. Guidelines, *NoCaseChange*{*https://https://help.glassdoor.com/article/Community-Guidelines/en_US*}, Accessed: 2019-03-21, 2019.

[243] J. J. Gumperz and D. H. Hymes, *Directions in sociolinguistics: The ethnography of communication*. B. Blackwell, 1986.

[244] E. Goffman, *Forms of talk*. University of Pennsylvania Press, 1981.

[245] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. EMNLP*, 2014.

[246] Glassdoor, "Glassdoor statistics," *expandedramblings.com/index.php/numbers-15-interesting-glassdoor-statistics*, 2018, Accessed: 2019-03-22.

[247] Glassdoor, "Glassdoor fraudulent reviews," *help.glassdoor.com/article/Fraudulent-reviews/en_US/*, 2019, Accessed: 2019-04-03.

[248] A. Chamberlain, "Give to get: A mechanism to reduce bias in online reviews," Glassdoor Research Report, October 2017, Tech. Rep., 2017.

[249] M. Goos, A. Manning, and A. Salomons, "Explaining job polarization: The roles of technology, offshoring and institutions," *Offshoring and Institutions (December 1, 2011)*, 2011.

[250] H. Heidemeier and K. Moser, "Self–other agreement in job performance ratings: A meta-analytic test of a process model.," *J. Appl. Psychol.*, 2009.

[251] P. Tambe and L. M. Hitt, "Now it's personal: Offshoring and the shifting skill composition of the us information technology workforce," *Management Science*, vol. 58, no. 4, pp. 678–695, 2012.

[252] R. A. Cooke and J. L. Szumal, "Measuring normative beliefs and shared behavioral expectations in organizations: The reliability and validity of the organizational culture inventory," *Psychological reports*, vol. 72, no. 3_suppl, pp. 1299–1330, 1993.

[253] D. Rousseau, "Quantitative assessment of organizational culture: The case for multiple," 1990.

[254] E. Fast, B. Chen, and M. S. Bernstein, "Empath: Understanding topic signals in large-scale text," in *Proc. CHI*, 2016.

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

[256] K. Saha, M. D. Reddy, and M. De Choudhury, "Joblex: A lexico-semantic knowledgebase of occupational information descriptors," in *SocInfo*, 2019.

[257] Fortune, "Fortune 500," *fortune.com/fortune500/list*, 2019, Accessed: 2019-03-27.

[258] Fortune, "Fortune 100 best companies," *fortune.com/best-companies/list/*, 2019, Accessed: 2019-03-27.

[259] R. Ingrassia, "Appearance or reality? monitoring of employer branding in public network space: The glassdoor case," 2017.

[260] T. Baldwin, P. Cook, M. Lui, A. MacKinlay, and L. Wang, "How noisy social media text, how diffrnt social media sources?" In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 2013, pp. 356–364.

[261] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman, "Social data: Biases, methodological pitfalls, and ethical boundaries," 2016.

[262] U. N. Pareek, *Training instruments for human resource development*. Tata McGraw-Hill, 1997.

[263] R. Eisenberger, P. Fasolo, and V. Davis-LaMastro, "Perceived organizational support and employee diligence, commitment, and innovation.," *Journal of applied psychology*, 1990.

[264] M. A. Campion, E. M. Papper, and G. J. Medsker, "Relations between work team characteristics and effectiveness: A replication and extension," *Personnel psychology*, 1996.

[265] J. V. Johnson and J. Lipscomb, "Long working hours, occupational health and the changing nature of work organization," *American journal of industrial medicine*, vol. 49, no. 11, pp. 921–929, 2006.

[266] K. Sparks, C. Cooper, Y. Fried, and A. Shirom, "The effects of hours of work on health: A meta-analytic review," *Journal of occupational and organizational psychology*, vol. 70, no. 4, pp. 391–408, 1997.

[267] C. C. Caruso *et al.*, "Long working hours, safety, and health: Toward a national research agenda," *American journal of industrial medicine*, vol. 49, no. 11, pp. 930–942, 2006.

[268] W. C. Shipley, *Shipley-2: manual*. WPS, 2009.

[269] R. Deshpandé, J. U. Farley, and F. E. Webster Jr, "Corporate culture, customer orientation, and innovativeness in japanese firms: A quadrad analysis," *Journal of marketing*, 1993.

[270] J. H. Greenhaus and S. Parasuraman, "Job performance attributions and career advancement prospects: An examination of gender and race effects," *Organizational Behavior and Human Decision Processes*, vol. 55, no. 2, pp. 273–297, 1993.

[271] M. A. Quińones, J. K. Ford, and M. S. Teachout, "The relationship between work experience and job performance: A conceptual and meta-analytic review," *Personnel psychology*, 1995.

[272] F. L. Schmidt, J. E. Hunter, and A. N. Outerbridge, "Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance.," *Journal of applied psychology*, 1986.

[273] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[274] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *Proc. SIGKDD*, ACM, 2008, pp. 7–15.

[275] K. Saha *et al.*, "Imputing Missing Social Media Data Stream in Multisensor Studies of Human Behavior," in *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII 2019)*, 2019.

[276] U. Grömping, "Variable importance assessment in regression: Linear regression versus random forest," *The American Statistician*, 2009.

[277] T. N. Y. Times, *Tracking coronavirus cases at u.s. colleges and universities*, Available at: https://www.nytimes.com/interactive/2021/us/college-covid-tracker.html, 2021.

[278] M. Korn, *Colleges begin mapping out a more normal fall—with caveats*, Available at: https://www.wsj.com/articles/colleges-begin-mapping-out-a-more-normal-fallwith-caveats-11615800600?reflink=desktopwebshare_permalink, 2021.

[279] R. C. Rabin, *C.d.c. officials say most available evidence indicates schools can be safe if precautions are taken on campus and in the community*, Available at: https://www.nytimes.com/2021/01/26/world/cdc-schools-reopening.html, 2021.

[280] A. Nierenberg and A. Pasick, *Schools briefing: University outbreaks and parental angst*, Available at: https://www.nytimes.com/2020/08/19/us/colleges-closing-covid.html, 2020.

[281] R. Rojas and M. Delkic, *As states reopen, governors balance existing risks with new ones*, Available at: https://www.nytimes.com/2020/05/17/us/coronavirus-states-reopen.html, 2020.

[282] E. S. Gurley, *Strategies to support the covid-19 response in lmics*, Accessed: 2021-02-03, 2020.

[283] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22 020–22 025, 2010.

[284] A. Wesolowski *et al.*, "Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data," *Proceedings of the National Academy of Sciences*, vol. 112, no. 35, pp. 11 114–11 119, 2015.

[285] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner, "Association between mobility patterns and covid-19 transmission in the usa: A mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 11, pp. 1247–1254, 2020.

[286] A. Rodríguez *et al.*, "Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence*, Apr. 2021.

[287] Y. Zhou, R. Xu, D. Hu, Y. Yue, Q. Li, and J. Xia, "Effects of human mobility restrictions on the spread of covid-19 in shenzhen, china: A modelling study using mobile phone data," *The Lancet Digital Health*, vol. 2, no. 8, e417–e424, 2020.

[288] S. Chang *et al.*, "Mobility network models of covid-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82–87, 2021.

[289] S. G. Benzell, A. Collis, and C. Nicolaides, "Rationing social contact during the covid-19 pandemic: Transmission risk and social benefits of us locations," *Proceedings of the National Academy of Sciences*, vol. 117, no. 26, pp. 14 642–14 644, 2020.

[290] Z. Mehrab *et al.*, "High resolution proximity statistics as early warning for us universities reopening during covid-19," *medRxiv*, 2020.

[291] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, "Population flow drives spatio-temporal distribution of covid-19 in china," *Nature*, vol. 582, no. 7812, pp. 389–394, 2020.

[292] C. O. Buckee *et al.*, "Aggregated mobility data could help fight covid-19," *Science (New York, NY)*, vol. 368, no. 6487, pp. 145–146, 2020.

[293] H. L. Hambridge, R. Kahn, and J.-P. Onnela, "Examining sars-cov-2 interventions in residential colleges using an empirical network," *International Journal of Infectious Diseases*, vol. 113, pp. 325–330, 2021.

[294] P. T. Gressman and J. R. Peck, "Simulating covid-19 in a university environment," *Mathematical biosciences*, vol. 328, p. 108 436, 2020.

[295] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," Stanford InfoLab, Tech. Rep., 1999.

[296] P. Bonacich, "Some unique properties of eigenvector centrality," *Social networks*, vol. 29, no. 4, pp. 555–564, 2007.

[297] M. E. Newman, "Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality," *Physical review E*, vol. 64, no. 1, p. 016 132, 2001.

[298] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[299] D. Mangrum and P. Niekamp, "Jue insight: College student travel contributed to local covid-19 spread," *Journal of Urban Economics*, p. 103 311, 2020.

[300] G. D. of Public Health, *Georgia department of public health daily status report*, Available at: https://dph.georgia.gov/covid-19-daily-status-report, 2020.

[301] G. Gibson *et al.*, "Surveillance-to-diagnostic testing program for asymptomatic sars-cov-2 infections on a large, urban campus in fall 2020," *Epidemiology*, vol. 33, no. 2, pp. 209–216, 2021.

[302] P. Keskinocak, B. E. Oruc, A. Baxter, J. Asplund, and N. Serban, "The impact of social distancing on covid19 spread: State of georgia case study," *Plos one*, vol. 15, no. 10, e0239798, 2020.

[303] K. Gaythorpe *et al.*, *Report 8: Symptom progression of covid-19*, 2020.

[304] B. Lopman *et al.*, "A modeling study to inform screening and testing interventions for the control of sars-cov-2 on university campuses," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[305] G. I. of Technology, *Georgia tech launches campus coronavirus testing*, Available at: https://health.gatech.edu/coronavirus/testing-launched, 2020.

[306] K. I. McKinnon, "Convergence of the nelder–mead simplex method to a nonstationary point," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 148–158, 1998.

[307] M. Andersen, "Early evidence on social distancing in response to covid-19 in the united states," *Available at SSRN 3569368*, 2020.

[308] U. of Illinois at Urbana-Champaign, *On-campus covid-19 testing*, Available at: https://covid19.illinois.edu/on-campus-covid-19-testing-data-dashboard/, 2020.

[309] K. A. Weeden and B. Cornwell, "The small-world network of college classes: Implications for epidemic spread on a university campus," *Sociological science*, vol. 7, pp. 222–241, 2020.

[310] M. Borowiak, F. Ning, J. Pei, S. Zhao, H.-R. Tung, and R. Durrett, "Controlling the spread of covid-19 on college campuses," *Mathematical biosciences and engineering: MBE*, vol. 18, no. 1, pp. 551–563, 2020.

[311] K. Kondo, "Simulating the impacts of interregional mobility restriction on the spatial spread of covid-19 in japan," *Scientific reports*, vol. 11, no. 1, p. 18 951, 2021.

[312] R. Bahl *et al.*, "Modeling covid-19 spread in small colleges," *Plos one*, vol. 16, no. 8, e0255654, 2021.

[313] L. Y. Saltzman, T. C. Hansel, and P. S. Bordnick, "Loneliness, isolation, and social support factors in post-covid-19 mental health.," *Psychological Trauma: Theory, Research, Practice, and Policy*, 2020.

[314] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[315] M. Harris and K. Holley, "Universities as anchor institutions: Economic and social potential for urban development," in *Higher education: Handbook of theory and research*, Springer, 2016, pp. 393–439.

[316] S. Watson, S. Hubler, D. Ivory, and R. Gebeloff, *A new front in america's pandemic: College towns*, Available at: https://www.nytimes.com/2020/09/06/us/colleges-coronavirus-students.html, 2020.

[317] D. Dave, A. Friedson, K. Matsuzawa, J. J. Sabia, and S. Safford, "Jue insight: Were urban cowboys enough to control covid-19? local shelter-in-place orders and coronavirus case growth," *Journal of urban economics*, p. 103 294, 2020.

[318] M. E. Loades *et al.*, "Rapid systematic review: The impact of social isolation and loneliness on the mental health of children and adolescents in the context of covid-19," *Journal of the American Academy of Child & Adolescent Psychiatry*, 2020.

[319] N. Ferguson *et al.*, "Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand,"

[320] S. S. Morse, R. L. Garwin, and P. J. Olsiewski, "Next flu pandemic: What to do until the vaccine arrives?," 2006.

[321] H. M. Weiss, "Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences," *Human resource management review*, vol. 12, no. 2, pp. 173–194, 2002.

[322] R. M. Groves and E. Peytcheva, "The impact of nonresponse rates on nonresponse bias: A meta-analysis," *Public opinion quarterly*, vol. 72, no. 2, pp. 167–189, 2008.

[323] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 1349–1380, 2000.

[324] L. Bucciarelli, *Engineering philosophy*. DUP Satellite; an imprint of Delft University Press, 2003.

[325] A. Muaremi, B. Arnrich, and G. Tröster, "Towards measuring stress with smartphones and wearable devices during workday and sleep," *BioNanoScience*, vol. 3, no. 2, pp. 172–183, 2013.

[326] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, IEEE, 2013, pp. 671–676.

[327] M. Ciman and K. Wac, "Individuals' stress assessment using human-smartphone interaction analysis," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 51–65, 2016.

[328] S. Amir, G. Coppersmith, P. Carvalho, M. J. Silva, and B. C. Wallace, "Quantifying mental health from social media with neural user embeddings," in *Machine Learning for Healthcare Conference*, PMLR, 2017, pp. 306–321.

[329] D. A. Weinberger, G. E. Schwartz, and R. J. Davidson, "Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress.," *Journal of abnormal psychology*, vol. 88, no. 4, p. 369, 1979.

[330] J. Hellhammer and M. Schubert, "The physiological response to trier social stress test relates to subjective measures of stress during but not before or after the test," *Psychoneuroendocrinology*, vol. 37, no. 1, pp. 119–124, 2012.

[331] T. A. Wearne *et al.*, "Anxiety sensitivity moderates the subjective experience but not the physiological response to psychosocial stress," *International Journal of Psychophysiology*, vol. 141, pp. 76–83, 2019.

[332] G. E. Silva *et al.*, "Relationship between reported and measured sleep times: The sleep heart health study (shhs)," *Journal of Clinical Sleep Medicine*, vol. 3, no. 06, pp. 622–630, 2007.

[333] S. K. Ernala *et al.*, "Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[334] S. Das and A. Kramer, "Self-censorship on facebook," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[335] K. K. Murdock, "Texting while stressed: Implications for students' burnout, sleep, and well-being.," *Psychology of Popular Media Culture*, vol. 2, no. 4, p. 207, 2013.

[336] U. Rimmele, R. Seiler, B. Marti, P. H. Wirtz, U. Ehlert, and M. Heinrichs, "The level of physical activity affects adrenal and cardiovascular reactivity to psychosocial stress," *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 190–198, 2009.

[337] A. Sadeh, G. Keinan, and K. Daon, "Effects of stress on sleep: The moderating role of coping style.," *Health Psychology*, vol. 23, no. 5, p. 542, 2004.

[338] S. K. Ernala *et al.*, "Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, 2019, p. 134.

[339] N. K. Denzin, "Triangulation 2.0," *Journal of mixed methods research*, vol. 6, no. 2, pp. 80–88, 2012.

[340] R. Wang *et al.*, "Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 886–897.

[341] S. K. Ernala, A. F. Rizvi, M. L. Birnbaum, J. M. Kane, and M. De Choudhury, "Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, p. 43, 2017.

[342] M. Sharmin *et al.*, "Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2015, pp. 505–516.

[343] S. Voida, D. J. Patterson, and S. N. Patel, "Sensor data streams," in *Ways of Knowing in HCI*, Springer, 2014, pp. 291–321.

[344] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. Springer, 2012.

[345] A. Liaw, M. Wiener, *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[346] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 2008.

[347] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: Extreme gradient boosting," *R package version 0.4-2*, pp. 1–4, 2015.

[348] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[349] G. Van Dijck and M. M. Van Hulle, “Speeding up the wrapper feature subset selection in regression by mutual information relevance and redundancy analysis,” in *International Conference on Artificial Neural Networks*, Springer, 2006, pp. 31–40.

[350] A. Airola, T. Pahikkala, W. Waegeman, B. De Baets, and T. Salakoski, “An experimental comparison of cross-validation techniques for estimating the area under the roc curve,” *Computational Statistics & Data Analysis*, vol. 55, no. 4, pp. 1828–1844, 2011.

[351] E. Goffman *et al.*, *The presentation of self in everyday life*. Harmondsworth London, 1978.

[352] J. A. Johnson, “The” self-disclosure” and” self-presentation” views of item response dynamics and personality scale validity.,” *Journal of Personality and Social Psychology*, vol. 40, no. 4, p. 761, 1981.

[353] M. M. Spada, A. V. Nikčević, G. B. Moneta, and A. Wells, “Metacognition, perceived stress, and negative emotion,” *Personality and Individual Differences*, vol. 44, no. 5, pp. 1172–1181, 2008.

[354] T. M. T. Do and D. Gatica-Perez, “Where and what: Using smartphones to predict next locations and applications in daily life,” *Pervasive and Mobile Computing*, vol. 12, pp. 79–91, 2014.

[355] F. A. Chowdhury *et al.*, “Modeling cyclic and ephemeral user behavior on social platforms,” in *Proceedings of the International AAAI Conference on Web and Social Media*, 2021.

[356] G. M. Harari, N. D. Lane, R. Wang, B. S. Crosier, A. T. Campbell, and S. D. Gosling, “Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges,” *Perspectives on Psychological Science*, vol. 11, no. 6, pp. 838–854, 2016.

[357] L. Chan, V. D. Swain, C. Kelley, K. de Barbaro, G. D. Abowd, and L. Wilcox, “Students’ experiences with ecological momentary assessment tools to report on emotional well-being,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 3, 2018.

[358] T. Plötz, “Applying machine learning for sensor data analysis in interactive systems: Common pitfalls of pragmatic use and ways to avoid them,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–25, 2021.

[359] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "Cstress: Towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, ACM, 2015, pp. 493–504.

[360] H. Sarker *et al.*, "Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, ACM, 2016, pp. 4489–4501.

[361] M. J. Gregoski, A. Vertegel, A. Shaporev, and F. A. Treiber, "Tension tamer: Delivering meditation with objective heart rate acquisition for adherence monitoring using a smart phone platform," *The Journal of Alternative and Complementary Medicine*, vol. 19, no. 1, pp. 17–19, 2013.

[362] A. Ghandeharioun and R. Picard, "Brightbeat: Effortlessly influencing breathing for cultivating calmness and focus," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 2017, pp. 1624–1631.

[363] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.

[364] X. Xu *et al.*, "Leveraging routine behavior and contextually-filtered features for depression detection among college students," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–33, 2019.

[365] M. B. Morshed *et al.*, "Prediction of mood instability with passive sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, p. 75, 2019.

[366] J. Rooksby, A. Morrison, and D. Murray-Rust, "Student perspectives on digital phenotyping: The acceptability of using smartphone data to assess mental health," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[367] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, communication & society*, vol. 15, no. 5, pp. 662–679, 2012.

[368] D. Lazer and J. Radford, "Data ex machina: Introduction to big data," *Annual Review of Sociology*, vol. 43, pp. 19–39, 2017.

[369] D. Estrin, "Small data, where n= me," *Communications of the ACM*, vol. 57, no. 4, pp. 32–34, 2014.

[370] P. Domingos, "Occam's two razors: The sharp and the blunt," in *KDD*, 1998, pp. 37–43.

[371] A. Mehrotra and M. Musolesi, "Using autoencoders to automatically extract mobility features for predicting depressive states," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, p. 127, 2018.

[372] A. H. Maslow, "The psychology of science a reconnaissance," 1966.

[373] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, p. 94, 2014.

[374] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: Pitfalls and promises," *Journal of autism and developmental disorders*, vol. 45, no. 5, pp. 1121–1136, 2015.

[375] L. Moldon, M. Strohmaier, and J. Wachs, "How gamification affects software developers: Cautionary evidence from a natural experiment on github," in *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, IEEE, 2021, pp. 549–561.

[376] M. Ortu, G. Destefanis, B. Adams, A. Murgia, M. Marchesi, and R. Tonelli, "The jira repository dataset: Understanding social aspects of software development," in *Proceedings of the 11th international conference on predictive models and data analytics in software engineering*, 2015, pp. 1–4.

[377] A. Thieme, D. Belgrave, and G. Doherty, "Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 27, no. 5, pp. 1–53, 2020.

[378] V. Das Swain, K. Saha, G. D. Abowd, and M. De Choudhury, "Social media and ubiquitous technologies for remote worker wellbeing and productivity in a post-pandemic world," in *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*, IEEE, 2020, pp. 121–130.

[379] M. Barbaro, *The rise of the workplace surveillance*, Aug. 24, 2022.

[380] G. Sewell and J. R. Barker, "Coercion versus care: Using irony to make sense of organizational surveillance," *Academy of Management review*, vol. 31, no. 4, pp. 934–961, 2006.

[381] H. Park, D. Ahn, K. Hosanagar, and J. Lee, “Human-ai interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–15.

[382] K. C. Kellogg, M. A. Valentine, and A. Christin, “Algorithms at work: The new contested terrain of control,” *Academy of Management Annals*, vol. 14, no. 1, pp. 366–410, 2020.

[383] K. Ball, “Workplace surveillance: An overview,” *Labor History*, vol. 51, no. 1, pp. 87–106, 2010.

[384] M. B. Calas and L. Smircich, “Past postmodernism? reflections and tentative directions,” *Academy of management review*, vol. 24, no. 4, pp. 649–672, 1999.

[385] T. B. Lawrence, M. I. Winn, and P. D. Jennings, “The temporal dynamics of institutionalization,” *Academy of management review*, vol. 26, no. 4, pp. 624–644, 2001.

[386] A. Aloisi and E. Gramano, “Artificial intelligence is watching you at work: Digital surveillance, employee monitoring, and regulatory issues in the eu context,” *Comp. Lab. L. & Pol’y J.*, vol. 41, p. 95, 2019.

[387] J. Kantor and A. Sundaram, *The rise of the worker productivity score*, Aug. 14, 2022.

[388] ActivTrak, \NoCaseChange{*https://www.activtrak.com/*}, Accessed: 2022-09-01, 2022.

[389] Humanyze, \NoCaseChange{*https://humanyze.com/*}, Accessed: 2022-09-01, 2022.

[390] K. Shilton, “Four billion little brothers? privacy, mobile phones, and ubiquitous data collection,” 2009.

[391] G. Grill and N. Andalibi, “Attitudes and folk theories of data subjects on transparency and accuracy in emotion recognition,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. CSCW1, pp. 1–35, 2022.

[392] R. Berger, “Now i see it, now i don’t: Researcher’s position and reflexivity in qualitative research,” *Qualitative research*, vol. 15, no. 2, pp. 219–234, 2015.

[393] S. Chancellor, E. P. Baumer, and M. De Choudhury, “Who is the” human” in human-centered machine learning: The case of predicting mental health from so-

cial media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–32, 2019.

[394] H. Nissenbaum, "Privacy as contextual integrity," *Wash. L. Rev.*, vol. 79, p. 119, 2004.

[395] M. S. Poole and A. H. Van de Ven, "Using paradox to build management and organization theories," *Academy of management review*, vol. 14, no. 4, pp. 562–578, 1989.

[396] D. A. Adler *et al.*, "Burnout and the quantified workplace: Tensions around personal sensing interventions for stress in resident physicians," 2022.

[397] S. L. Star and J. R. Griesemer, "Institutional ecology,translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39," *Social studies of science*, vol. 19, no. 3, pp. 387–420, 1989.

[398] A. Crabtree and R. Mortier, "Human data interaction: Historical lessons from social studies and cscw," in *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, Springer, 2015, pp. 3–21.

[399] I. Manokha, "The implications of digital employee monitoring and people analytics for power relations in the workplace," *Surveillance and Society*, vol. 18, no. 4, 2020.

[400] X. Jiang, J. I. Hong, and J. A. Landay, "Approximate information flows: Socially-based modeling of privacy in ubiquitous computing," in *International Conference on Ubiquitous Computing*, Springer, 2002, pp. 176–193.

[401] M. Abraham *et al.*, "Electronic monitoring at work: The role of attitudes, functions, and perceived control for the acceptance of tracking technologies," *Human Resource Management Journal*, vol. 29, no. 4, pp. 657–675, 2019.

[402] S. Chen, *Chinese construction firms using ai to monitor workers' safety ... but also to spot 'loiterers'*, 2021.

[403] RemoteDesk, *NoCaseChange*{*https://www.remotedesk.com/solutions/webcam-monitoring*}, Accessed: 2022-09-01, 2022.

[404] Interguard, *NoCaseChange*{*https://www.interguardsoftware.com/employee-monitoring-screenshots/*}, Accessed: 2022-09-01, 2022.

[405] F. System, *NoCaseChange*{*https://fmsystems.com/products/workplace-analytics/sensor-analytics/*}, Accessed: 2022-09-01, 2022.

[406] Freespace, \NoCaseChange{*https://www.afreespace.com/solutions/enabling-hybrid-working/#sensors*}, Accessed: 2022-09-01, 2022.

[407] V. Insights, \NoCaseChange{*https://techcommunity.microsoft.com/t5/microsoft-viva-blog/daily-briefing-and-myanalytics-branding-updates-to-reflect/ba-p/2681246*}, Accessed: 2022-09-01, 2022.

[408] M. Analytics, \NoCaseChange{*https://docs.microsoft.com/en-us/viva/insights/personal/use/dashboard-2*}, Accessed: 2022-09-01, 2022.

[409] K. Kelly and G. Wolf, "What is the quantified self," *The Quantified Self*, vol. 5, p. 2007, 2007.

[410] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, "Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach," in *9th IEEE International Workshop on Advanced Motion Control, 2006.*, IEEE, 2006, pp. 750–755.

[411] S. Davidoff, M. K. Lee, A. K. Dey, and J. Zimmerman, "Rapidly exploring application design through speed dating," in *International conference on ubiquitous computing*, Springer, 2007, pp. 429–446.

[412] V. Ikonen and K. Rentto, "Scenario evaluations for ubiquitous computing-stories come true?" In *Workshop on User-Centered Evaluation of Ubiquitous Computing Application*, 2002.

[413] L. Little, S. Marsh, and P. Briggs, "Trust and privacy permissions for an ambient world," in *Trust in e-services: Technologies, Practices and Challenges*, IGI Global, 2007, pp. 259–292.

[414] L. Little, E. Sillence, and P. Briggs, "Ubiquitous systems and the family: Thoughts about the networked home," in *Proceedings of the 5th Symposium on Usable Privacy and Security*, 2009, pp. 1–9.

[415] J. Newn, R. M. Kelly, S. D'Alfonso, and R. Lederman, "Examining and promoting explainable recommendations for personal sensing technology acceptance," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–27, 2022.

[416] M. K. Lee, "Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management," *Big Data & Society*, vol. 5, no. 1, p. 2 053 951 718 756 684, 2018.

[417] C. T. Wolf, "Explainability scenarios: Towards scenario-based xai design," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 2019, pp. 252–257.

[418] A. Tversky, "Features of similarity.," *Psychological review*, vol. 84, no. 4, p. 327, 1977.

[419] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[420] M. Langheinrich, "Privacy by design—principles of privacy-aware ubiquitous systems," in *International conference on ubiquitous computing*, Springer, 2001, pp. 273–291.

[421] J. Nicholas, K. Shilton, S. M. Schueller, E. L. Gray, M. J. Kwasny, D. C. Mohr, *et al.*, "The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: Cross-sectional questionnaire study," *JMIR mHealth and uHealth*, vol. 7, no. 4, e12578, 2019.

[422] C. Utz *et al.*, "Apps against the spread: Privacy implications and user acceptance of covid-19-related smartphone apps on three continents," in *Proceedings of the 2021 chi conference on human factors in computing systems*, 2021, pp. 1–22.

[423] C. Röcker, "Acceptance of future workplace systems: How the social situation influences the usage intention of ambient intelligence technologies in work environments," in *Proceedings of the 9th International Conference on Work With Computer Systems*, 2009, pp. 9–14.

[424] D. Tran, H. M. Do, W. Sheng, H. Bai, and G. Chowdhary, "Real-time detection of distracted driving based on deep learning," *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1210–1219, 2018.

[425] P. Moore and A. Robinson, "The quantified self: What counts in the neoliberal workplace," *new media & society*, vol. 18, no. 11, pp. 2774–2792, 2016.

[426] D. E. Sappington, "Incentives in principal-agent relationships," *Journal of economic Perspectives*, vol. 5, no. 2, pp. 45–66, 1991.

[427] J. G. Adair, "The hawthorne effect: A reconsideration of the methodological artifact.," *Journal of applied psychology*, vol. 69, no. 2, p. 334, 1984.

[428] V. Das Swain, M. D. Reddy, K. A. Nies, L. Tay, M. De Choudhury, and G. D. Abowd, "Birds of a feather clock together: A study of person-organization fit through latent activity routines," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–30, 2019.

[429] C. Zakaria, R. Balan, and Y. Lee, “Stressmon: Scalable detection of perceived stress and depression using passive sensing of changes in work routines and group interactions,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–29, 2019.

[430] H. Xu, H.-H. Teo, B. C. Tan, and R. Agarwal, “The role of push-pull technology in privacy calculus: The case of location-based services,” *Journal of management information systems*, vol. 26, no. 3, pp. 135–174, 2009.

[431] J. Grudin, “Why cscw applications fail: Problems in the design and evaluationof organizational interfaces,” in *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, 1988, pp. 85–93.

[432] I. Ajunwa, K. Crawford, and J. Schultz, “Limitless worker surveillance,” *Calif. L. Rev.*, vol. 105, p. 735, 2017.

[433] J. Finch, “The vignette technique in survey research,” *Sociology*, vol. 21, no. 1, pp. 105–114, 1987.

[434] H. Aguinis and K. J. Bradley, “Best practice recommendations for designing and implementing experimental vignette methodology studies,” *Organizational research methods*, vol. 17, no. 4, pp. 351–371, 2014.

[435] R. Ludwick and R. A. Zeller, “The factorial survey: An experimental method to replicate real world problems,” *Nursing research*, vol. 50, no. 2, pp. 129–133, 2001.

[436] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” *MIS quarterly*, pp. 319–340, 1989.

[437] W. R. King and J. He, “A meta-analysis of the technology acceptance model,” *Information & management*, vol. 43, no. 6, pp. 740–755, 2006.

[438] D. Kim, K. Park, Y. Park, and J.-H. Ahn, “Willingness to provide personal information: Perspective of privacy calculus in iot services,” *Computers in Human Behavior*, vol. 92, pp. 273–281, 2019.

[439] É. Charbonneau and C. Doberstein, “An empirical assessment of the intrusiveness and reasonableness of emerging work surveillance technologies in the public sector,” *Public Administration Review*, vol. 80, no. 5, pp. 780–791, 2020.

[440] J. Cheney-Lippold, *We are data*. New York University Press, 2017.

[441] J. S. Seberger and S. Patil, “Us and them (and it): Social orientation, privacy concerns, and expected use of pandemic-tracking apps in the united states,” in *Pro-*

ceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–19.

[442] Y. Sun, N. Wang, X.-L. Shen, and J. X. Zhang, “Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences,” *Computers in Human Behavior*, vol. 52, pp. 278–292, 2015.

[443] J. Sheringham, I. Kuhn, and J. Burt, “The use of experimental vignette studies to identify drivers of variations in the delivery of health care: A scoping review,” *BMC medical research methodology*, vol. 21, no. 1, p. 81, 2021.

[444] E.-M. Zeissig, C. Lidynia, L. Vervier, A. Gadeib, and M. Ziefle, “Online privacy perceptions of older adults,” in *Human Aspects of IT for the Aged Population. Applications, Services and Contexts: Third International Conference, ITAP 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 3*, Springer, 2017, pp. 181–200.

[445] S. S. Tzafrir and S. L. Dolan, “Trust me: A scale for measuring manager-employee trust,” *Management Research: Journal of the Iberoamerican Academy of Management*, 2004.

[446] R. J. Karren and M. W. Barringer, “A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice,” *Organizational Research Methods*, vol. 5, no. 4, pp. 337–361, 2002.

[447] P. De Boeck *et al.*, “The estimation of item response models with the lmer function from the lme4 package in r,” *Journal of Statistical Software*, vol. 39, pp. 1–28, 2011.

[448] P. E. Shrout and N. Bolger, “Mediation in experimental and nonexperimental studies: New procedures and recommendations.,” *Psychological methods*, vol. 7, no. 4, p. 422, 2002.

[449] D. Tingley, T. Yamamoto, K. Hirose, L. Keele, and K. Imai, “Mediation: R package for causal mediation analysis,” 2014.

[450] U. Ehsan *et al.*, “The who in explainable ai: How ai background shapes perceptions of ai explanations,” *arXiv preprint arXiv:2107.13509*, 2021.

[451] J. M. Keynes, “Economic possibilities for our grandchildren,” in *Essays in persuasion*, Springer, 1930, pp. 321–332.

[452] L. Elliot, “Economics: Whatever happened to keynes’ 15-hour working week?” *https://www.theguardian.com/business/2008/sep/01/economics*, 2008.

[453] J. Harter, "Is quiet quitting real?" *https://www.gallup.com/workplace/398306/quiet-quitting-real.aspx*, 2023.

[454] J. Clifton, "The real global jobs crisis," *https://www.gallup.com/analytics/471212/real-global-jobs-crisis.aspx*, 2022.

[455] J. Suzman, *Work: A history of how we spend our time*. Bloomsbury Publishing, 2020.

[456] C. D. Kidd *et al.*, "The aware home: A living laboratory for ubiquitous computing research," in *Cooperative Buildings. Integrating Information, Organizations, and Architecture: Second International Workshop, CoBuild'99, Pittsburgh, PA, USA, October 1-2, 1999. Proceedings 2*, Springer, 1999, pp. 191–198.

[457] S. S. Intille, K. Larson, J. Beaudin, J. Nawyn, E. M. Tapia, and P. Kaushik, "A living laboratory for the design and evaluation of ubiquitous computing technologies," in *CHI'05 extended abstracts on Human factors in computing systems*, 2005, pp. 1941–1944.

[458] S. Papagiannidis and D. Marikyan, "Smart offices: A productivity and well-being perspective," *International Journal of Information Management*, vol. 51, p. 102 027, 2020.

[459] S. Boovaraghavan *et al.*, "Mites: Design and deployment of a general-purpose sensing infrastructure for buildings," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1–32, 2023.

[460] E. Guo and T. Ryan-Mosley, "Computer scientists designing the future can't agree on what privacy means," *https://www.technologyreview.com/2023/04/03/1070665/cmu-university-privacy-battle-smart-building-sensors-mites/*, 2023.

[461] E. Margariti, V. Vlachokyriakos, and D. Kirk, "Understanding occupants' experiences in quantified buildings: Results from a series of exploratory studies.," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–15.

[462] V. Das Swain, S. Williams, A. Fourney, and S. T. Iqbal, "Two birds with one phone: The role of mobile use in the daily practices of remote information work," in *2022 Symposium on Human-Computer Interaction for Work*, 2022, pp. 1–8.

[463] R. K. Schwer, M. C. Mejza, and M. Grun-Rehomme, "Workplace violence and stress: The case of taxi drivers," *Transportation Journal*, vol. 49, no. 2, pp. 5–23, 2010.

[464] O. O. Ogancha, “Power asymmetry and the quest for inclusiveness in the workplace,” *American Economic Review*, vol. 57, p. 62, 2019.

[465] J. Cho and S. Voida, “Envisioning new productivity tools for domestic information work environments,” 2020.

[466] V. Das Swain *et al.*, “Focused time saves nine: Evaluating computer–assisted protected time for hybrid information work,” in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–18.

[467] H. Kaur, D. McDuff, A. C. Williams, J. Teevan, and S. T. Iqbal, ““i didn’t know i looked angry”: Characterizing observed emotion and reported affect at work,” in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–18.

[468] J. Rawls, “A theory of justice,” in *Ethics*, Routledge, 2004, pp. 229–234.

[469] P.-Y. Hung and M. S. Ackerman, “Helping people to control their everyday data for care: A scenario-based study,” in *International Conference on Pervasive Computing Technologies for Healthcare*, Springer, 2021, pp. 272–301.

[470] D. Calacci, “Organizing in the end of employment: Information sharing, data stewardship, and digital workerism,” in *2022 Symposium on Human-Computer Interaction for Work*, 2022, pp. 1–9.

[471] N. Newman, “Reengineering workplace bargaining: How big data drives lower wages and how reframing labor law can restore information equality in the workplace,” *U. Cin. L. Rev.*, vol. 85, p. 693, 2017.

[472] V. Khovanskaya, L. Dombrowski, J. Rzeszotarski, and P. Sengers, “The tools of management: Adapting historical union tactics to platform-mediated labor,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–22, 2019.

[473] K. Gregory, “’worker data science’ can teach us how to fix the gig economy,” *https://www.wired.com/story/labor-organizing-unions-worker-algorithms/*, 2021.

[474] N. Salehi, L. C. Irani, M. S. Bernstein, A. Alkhatib, E. Ogbe, and K. Milland, “We are dynamo: Overcoming stalling and friction in collective action for crowd workers,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 1621–1630.

[475] L. D. Introna, “Workplace surveillance, privacy and distributive justice,” *Acm Sigcas Computers and Society*, vol. 30, no. 4, pp. 33–39, 2000.

[476] A. Chaudhry *et al.*, “Personal data: Thinking inside the box,” 2015.

[477] A. Crabtree and R. Mortier, “Personal data, privacy and the internet of things: The shifting locus of agency and control,” *Privacy and the Internet of Things: The Shifting Locus of Agency and Control (November 22, 2016)*, 2016.

[478] E. Kaplan, “The spy who fired me: The human costs of workplace monitoring,” *Harper’s Magazine*, vol. 1135, 2015.

[479] K. Bednar, S. Spiekermann, and M. Langheinrich, “Engineering privacy by design: Are engineers ready to live up to the challenge?” *The Information Society*, vol. 35, no. 3, pp. 122–142, 2019.

[480] S. Chowdhary, A. Kawakami, M. L. Gray, J. Suh, A. Olteanu, and K. Saha, “Can workers meaningfully consent to workplace wellbeing technologies?” In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023, pp. 569–582.

[481] A. Alesina, E. Glaeser, and B. Sacerdote, “Work and leisure in the united states and europe: Why so different?” *NBER macroeconomics annual*, vol. 20, pp. 1–64, 2005.

[482] P. M. Secunda, “The employee right to disconnect,” *Notre Dame J. Int’l Comp. L.*, vol. 9, p. 1, 2019.

[483] J. Kelly, “Belgium, portugal and other european countries prohibit managers from contacting employees outside of working hours,” *https://www.forbes.com/sites/jackkelly/2022/02/03/belgium-portugal-and-other-european-countries-are-ahead-of-the-us-prohibiting-managers-from-contacting-employees-outside-of-working-hours/?sh=546baf3b1d00*, 2022.

[484] B. Van Ark, M. O’Mahony, and M. P. Timmer, “The productivity gap between europe and the united states: Trends and causes,” *Journal of economic perspectives*, vol. 22, no. 1, pp. 25–44, 2008.

[485] J. Clifton and P. Sinyan, “Europe gets life right, but work wrong,” *https://www.gallup.com/workplace/393794/europe-gets-life-right-work-wrong.aspx*, 2022.

[486] P. Regulation, “General data protection regulation,” *Intouch*, vol. 25, pp. 1–5, 2018.

[487] X. W. Chan, S. Shang, P. Brough, A. Wilkinson, and C.-q. Lu, “Work, life and covid-19: A rapid review and practical recommendations for the post-pandemic workplace,” *Asia Pacific journal of human resources*, vol. 61, no. 2, pp. 257–276, 2023.

[488] A. Sutton and C. Atkinson, “Performance at the cost of well-being? testing the multi-level effects of hr practices on organisational performance via employee ex-

periences and well-being," in *Evidence-based HRM: a Global Forum for Empirical Scholarship*, Emerald Publishing Limited, 2023.

[489] C. Cregan, C. T. Kulik, S. Johnston, and T. Bartram, "The influence of calculative ("hard") and collaborative ("soft") hrm on the layoff-performance relationship in high performance workplaces," *Human Resource Management Journal*, vol. 31, no. 1, pp. 202–224, 2021.

[490] D. Allhutter, F. Cech, F. Fischer, G. Grill, and A. Mager, "Algorithmic profiling of job seekers in austria: How austerity politics are made effective," *frontiers in Big Data*, p. 5, 2020.

[491] A. L. Kalleberg, "Precarious work, insecure workers: Employment relations in transition," *American sociological review*, vol. 74, no. 1, pp. 1–22, 2009.

[492] S. J. Gould, *Mismeasure of man*. WW Norton & company, 1996.

[493] I. Manokha, "Why the rise of wearable tech to monitor employees is worrying," *The Conversation*, vol. 3, 2017.

[494] A. Kawakami *et al.*, "Sensing wellbeing in the workplace, why and for whom? envisioning impacts with organizational stakeholders," *arXiv preprint arXiv:2303.06794*, 2023.

[495] G. D. Nord, T. F. McCubbins, and J. H. Nord, "E-monitoring in the workplace: Privacy, legislation, and surveillance software," *Communications of the ACM*, vol. 49, no. 8, pp. 72–77, 2006.

[496] E. P. I. Center, "Workplace privacy," *https://archive.epic.org/privacy/workplace/*, 2022.

[497] M. Mitchell *et al.*, "Model cards for model reporting," in *Proceedings of the conference on fairness, accountability, and transparency*, 2019, pp. 220–229.

[498] B. N. Waber, D. Olguin Olguin, T. Kim, and A. Pentland, "Productivity through coffee breaks: Changing social networks by changing break structure," *Available at SSRN 1586375*, 2010.

[499] S. Kapania, O. Siy, G. Clapper, A. M. SP, and N. Sambasivan, """ because ai is 100% right and safe": User attitudes and sources of ai authority in india," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–18.

[500] A. McIntyre, *Participatory action research*. Sage Publications, 2007.

[501] M. K. Lee *et al.*, “Webuildai: Participatory framework for algorithmic governance,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, no. CSCW, pp. 1–35, 2019.