A FRAMEWORK FOR
DOMAIN-DRIVEN DEVELOPMENT OF
PERSONAL HEALTH INFORMATICS
TECHNOLOGIES

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Elizabeth Lindley Murnane
January 2017
This dissertation advances a vision of Personal Health Informatics (PHI), a class of tools that can leverage personal data to support health self-management. Today, a powerful combination of factors is coming together that can facilitate the creation of these technologies and amplify their benefits. Namely, the world is awash in data, software and sensors continue to capture more, increasingly capable algorithms are helping humans make sense of it all, and ubiquitous devices (that people are keen to use to manage their wellness) can deliver this information via individually-tailored, insight-enabling, personally-empowering, health-enhancing feedback.

A central argument of this dissertation is that domain knowledge can help drive PHI development in order to fully capitalize on the potential of these technologies. A central contribution of this dissertation is a framework for engaging in domain-driven development. In specifying this reusable development pattern, I provide guidance on moving through stages of domain inquiry, domain-driven health assessment, and domain-aware intervention design.

To begin, I describe what domain knowledge encompasses, why it is valuable, and how to synthesize insights from diverse sources in order to gain an appreciation of the role technology can play in a given context. I then explain how this understanding can inform research goals, strategies for assessing significant
health determinants, and implications for designing effective interventions.

To demonstrate this process in practice, I present my own research as a case study on developing domain-driven technology that supports healthy sleep, daily performance, and emotional wellbeing. Overall, I argue that a domain-driven approach that foregrounds a deep understanding of a targeted aspect of health, together with a compassion for the lived experiences of users, will produce technological solutions that better meet individual needs and promote more positive outcomes.
BIOGRAPHICAL SKETCH

Broadly speaking, Elizabeth’s research interests lie in human-computer interaction, personal informatics, recommender systems, social computing, and personalization. She is particularly compelled by applications in the domains of personal information management, civic innovation, and health.

Elizabeth received her Bachelor of Science in Mathematics with Computer Science in 2007 from the Massachusetts Institute of Technology (MIT), where her undergraduate research in the Computer Science and Artificial Intelligence Laboratory (CSAIL) focused on information visualization and conversational agents. Upon graduating, Elizabeth co-founded and spent four years as the lead engineer of Architexa, a CSAIL spinoff that built interactive visualization tools to help software developers make sense of and share important aspects of source code.

Since 2011, Elizabeth has been a PhD student in Information Science at Cornell University. In her graduate studies, she has continued to explore her overarching research goal: designing and deploying technologies that are aware of idiosyncratic user needs and that support people in managing various aspects of their daily lives. Much of Elizabeth’s current work (out of which this dissertation grew) is on developing novel mediums for manually collecting personal data, lightweight algorithms for passive sensing and user modeling, and interfaces that provide tailored feedback and experiences.

In pursuing these directions, Elizabeth collaborates with interdisciplinary teams comprised of domain experts from critical theorists to legal scholars to clinicians. Through these partnerships she hopes to continue pushing the boundaries of how systems can enable more positive interactions with and through technology, on individual, group, and societal levels.
ACKNOWLEDGEMENTS

I would like to extend a huge thanks to everyone who supported me throughout the course of my PhD and helped to make this dissertation happen.

I am especially appreciative of my fantastic committee who saw me through. Dan Cosley has embodied what it means to be a true mentor, Geri Gay has been an incredible advocate and often felt like a co-advisor, and Claire Cardie has been a valuable source of interdisciplinary perspective.

Humbly, I realize that it would not be possible to acknowledge by name every other individual who has had a positive impact on me over the years; so though I give broad thanks, it is heartfelt. I am tremendously grateful for my family as well as friends, colleagues, and collaborators from Cornell University and beyond — I look forward to our continued connection. I am also indebted to all of the open-minded participants willing to be involved in my research and all the known or anonymous reviewers that helped shape its presentation.

Finally, I feel very fortunate to have received a Graduate Research Fellowship from the National Science Foundation, whose funding afforded me the flexibility to undertake a number of diverse projects, broaden my horizons, and transform ideas into reality.
# TABLE OF CONTENTS

Biographical Sketch ........................................ iii
Acknowledgements ........................................... iv
Table of Contents ........................................... v
List of Tables .................................................. vii
List of Figures ................................................. viii

1 Introduction ................................................. 1
   1.1 Motivation ............................................. 2
   1.2 Dissertation Overview & Contributions ................. 7

2 Background .................................................. 10
   2.1 The Roles of Technology in Supporting Health .......... 10
      2.1.1 The Roots of Modern Medicine ..................... 11
      2.1.2 The Age of Behavior Change ....................... 14
      2.1.3 The Individual as the Nexus of Health Management .. 19
   2.2 Personal Health Informatics (PHI) ......................... 21
      2.2.1 Capturing Personal Data .......................... 25
      2.2.2 Analyzing Data to Assess Health .................. 32
      2.2.3 Delivering User-Facing Feedback ................. 34
   2.3 Domain-Driven Personal Health Informatics ............... 48
      2.3.1 Defining Domain Knowledge ....................... 48
      2.3.2 The Value of Domain-Driven PHI ................... 49
      2.3.3 Examples of Domain-Driven PHI .................... 51
      2.3.4 A Domain–Practice Gap in HCI .................... 53
   2.4 A Framework for Domain-Driven PHI Development .......... 57

3 Domain Inquiry .............................................. 62
   3.1 An Overview of the Inquiry Process ....................... 63
   3.2 Selecting an Application Area .......................... 70
      3.2.1 Sleep, Cognitive Performance, and Emotional Wellness .. 70
      3.2.2 Opportunities for Technology ....................... 74
      3.2.3 Related Academic and Commercial Work ............. 75
      3.2.4 Connecting with Chronobiology ..................... 81
   3.3 Gathering Domain Knowledge ............................ 84
      3.3.1 Chronobiology & Circadian Rhythms ................ 85
      3.3.2 Chronotype ......................................... 87
      3.3.3 Circadian Disruption .............................. 88
      3.3.4 Traditional Assessment Methods from Chronobiology .. 91
   3.4 Informing PHI Development .............................. 94
      3.4.1 Defining Scope ................................... 94
      3.4.2 Soft Sensing for Health Assessment ............... 97
4  Domain-Driven Health Assessment  
4.1  Method .................................................. 100  
4.1.1  Participants & Procedures ......................... 101  
4.1.2  Data .................................................. 102  
4.2  Experiment 1 Findings ................................. 111  
4.2.1  Daily Rhythms in Sleep and Social Technology Use .. 111  
4.2.2  Leveraging Social Data for Sleep Sensing .......... 114  
4.2.3  Sleep’s Links with Neurobehavioral Functioning ...... 122  
4.3  Experiment 2 Findings ................................. 129  
4.3.1  Smartphone Application Usage Patterns ............. 129  
4.3.2  Circadian Alertness Rhythms Reflected in App Use ... 132  
4.3.3  Connecting App Use and Alertness with Sleep ........ 138  
4.4  Discussion .............................................. 140  
4.4.1  Limitations and Future Work ....................... 144  
4.4.2  Implications for Domain-Driven Health Assessment .. 147  

5  Domain-Aware Intervention Design .................... 151  
5.1  Chronobiology-Aware PHI .......................... 153  
5.2  Chronobiology-Informed Sleep Support ............... 154  
5.2.1  Personalizing Sleep Hygiene Recommendations ...... 155  
5.2.2  Stabilizing Circadian Disruptions ................... 158  
5.3  Biologically-Friendly Productivity Technology ......... 165  
5.3.1  Self- and Social-Awareness ........................ 166  
5.3.2  Scheduling and Activity Management ................. 171  
5.3.3  Performance-Predictive Systems ..................... 175  
5.4  Mental Health Management .......................... 178  
5.4.1  Participatory Design of MoodRhythm ................. 179  
5.4.2  Iteratively Refining Design Guidelines ............... 182  
5.4.3  Implications for Preemptive Interventions .......... 187  

6  General Discussion and Conclusion .................... 190  
6.1  Summary of Chapters and Contributions .............. 191  
6.2  Discussion .............................................. 194  
6.2.1  Technology as a Double-Edged Sword ............... 195  
6.2.2  Manual and Passive Modes ........................ 201  
6.2.3  Avoiding Over-Personalization of Tailored Experiences . 204  
6.2.4  A Dual (Not Duel) Domain & Data Driven Approach .. 210  
6.3  Opportunities for Future Work ....................... 215  
6.3.1  Moving From Personal to Collective Informatics .... 215  
6.3.2  Applying the Framework in Areas Beyond Health .... 218  
6.3.3  Open Challenges of Data, Analysis, and Design ...... 224  
6.4  Concluding Remarks ................................... 229  

Bibliography ................................................. 230
### LIST OF TABLES

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Categories of applications used by participants in Experiment 2, with examples and amounts of applications and usage events.</td>
<td>109</td>
</tr>
<tr>
<td>4.2</td>
<td>Average sleep duration for each participant according to social-sensor data, screen on/off data, and ground truth sleep diary data. (* denotes inferences that fall within 95% confidence interval based on diary self-reports, $p &lt; .01$).</td>
<td>114</td>
</tr>
<tr>
<td>4.3</td>
<td>Median values of CMC-based activity levels following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name ((*p &lt; .001, **p &lt; .0001)).</td>
<td>124</td>
</tr>
<tr>
<td>4.4</td>
<td>Median values of cognitive functioning variables following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name ((*p &lt; .05, **p &lt; .01)).</td>
<td>126</td>
</tr>
<tr>
<td>4.5</td>
<td>Median values of sentiment expressed in Facebook posts following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name ((**p &lt; .0001)).</td>
<td>127</td>
</tr>
<tr>
<td>4.6</td>
<td>Median values of usage features during low vs. high alertness. Significant differences in medians marked on variable name ((*p &lt; .05)).</td>
<td>137</td>
</tr>
<tr>
<td>5.1</td>
<td>Guidelines for designing self-monitoring technologies to support mental health management.</td>
<td>186</td>
</tr>
<tr>
<td>5.2</td>
<td>Variations in technology use identified as characteristic of mania and depression.</td>
<td>188</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.1 A framework for domain-driven PHI development, comprised of stages for tapping domain knowledge, collecting and analyzing personal data to assess health, and designing user-facing feedback. ................................................. 60

4.1 Chronotypes of participants in Experiment 1. ................. 104
4.2 Distribution of participant chronotypes in Experiment 2. .... 104
4.3 Daily trends in participants’ average CMC-based usage. ....... 104
4.4 Average sleep duration on work days and free days reveal the “scissors of sleep” phenomenon — a discrepant pattern of sleep on work days versus free days that is reversed for early and late chronotypes. .......................................................... 117
4.5 Socially-sensed average social jet lag (discrepancy between mid-sleep on free days and work days) across chronotypes. .... 118
4.6 Shifts in sleep midpoint across study phases. .................. 119
4.7 Sleep duration and sentiment the following day. ............... 127
4.8 Hourly smartphone app usage by category. .................... 130
4.9 Use across the week of productivity and entertainment apps shown with standard error. .......................................................... 131
4.10 Percentage increase (positive y value) or decrease (negative y value) in amount of usage by early types compared to usage by late types of productivity and entertainment apps across the day. 133
4.11 Temporal trends in application use (Usage) and alertness performance (Performance) across internal body clock time (InT). Usage axis is proportion (normalized to [0,1] scale) of all an app category’s usage events that occurred in a given hour. Performance axis is percent deviation from individual baseline of alertness measured in a given hour. Internal time axis is number of hours since biological midnight, and accompanying spectrum indicates periods of the biological day. ....................... 135

5.1 To provide peripheral self-awareness, this live wallpaper’s color transitions in real-time in accordance with the user’s alertness levels at that moment. By default, brighter, more yellow saturation corresponds to higher alertness while a faded blue-gray color corresponds to lower alertness. ....................... 167
5.2 This chronobiology-aware calendar background scaffolds self-awareness of personal alertness levels, which are represented using a customizable color scale. Visual indicators on events provide an at-a-glance sense of whether scheduling aligns with personal alertness at that time. ....................... 168
5.3 The color of personal “beacons” transitions over the course of the day to display personal alertness levels in a way visible to other people. Beacons could be objects placed in the environment (e.g., a desk ornament, lamp, etc.) or wearables worn on the body (e.g., a necklace, pin, bracelet, ring, etc).

5.4 This “family portrait” (or “roommate portrait”, etc.) interface syncs to personal devices and uses a shared digital display in the home to deliver chronobiology information (e.g., performance and sleep-wake patterns) about household members as a way to coordinate scheduling and improve interpersonal awareness and empathy.

5.5 Specifying performance-related parameters (e.g., alertness, as shown here) enables scheduling recommendations that align with personal rhythms.

5.6 This “Fix My Day” scheduling assistant analyzes the day’s events in order to provide suggestions for rearrangements that optimize alignment with personal performance patterns.

5.7 Smartphone and smartwatch clock widgets show glanceable views of personalized activity recommendations.

5.8 Storyboard of a playful chronobiology-aware activity recommender that provides in-the-moment feedback through an interactive, haptic experience.

5.9 This interactive sleep-decision-support tool provides feedback about the net gains or losses of staying awake (e.g., to study for a test) in terms of the predicted impact of specified sleep choices on performance the next day.

5.10 Screenshots of MoodRhythm.

5.11 Percentages of respondents (y axis) who report tracking specified indicators (x axis).

5.12 Percentages (y axis) of various methods (key) used to track specified indicator (x axis).

5.13 Example of one participant’s custom tracking setup that captures personally meaningful variables, assesses daily status, and delivers messages of encouragement. Diary entries have been blurred to protect the participant’s privacy.
CHAPTER 1
INTRODUCTION

The World Health Organization defines health as “a state of complete physical, mental, and social wellbeing — and not merely the absence of disease or infirmity” [533]. To achieve the full breadth of this definition, health care is breaking free from its traditional illness-centered, doctor-dependent model of visit–test–treat and moving toward a person-centric vision of medicine that is more proactive, personalized, and self-driven. Making the individual the nexus of her own health management, with a focus on overall wellness and prevention, could be pivotal in solving the root of contemporary public health challenges [475].

Technology can play a major role in this metamorphosis. Specifically, computing advances open new avenues to better understand and shape human behaviors — which is critical, given that a person’s behavioral and lifestyle choices currently provide the single greatest opportunity to improve health amidst today’s staggering prevalence of chronic disease [449]. Indeed, researchers within the Human-Computer Interaction (HCI) community and relevant sub-areas (e.g., personal informatics, persuasive computing, mobile health) are also identifying opportunities where interactive systems can play a positive, perhaps transformative, role in addressing modern society’s health problems.

Still, further steps must be taken in making these possibilities a reality. Most relevant to this dissertation is ensuring that tools are personally relevant, contextually appropriate, and lead to genuine health improvements. I argue this goal can be made more achievable by deeply grounding development decisions in knowledge from health-relevant domains.
1.1 Motivation

Today, lifestyle choices are the root of poor health for many people. These behavioral health determinants are linked to preventable, chronic diseases that require extended management and come with a heavy price, in terms of both lost quality of life as well as financial spending [465].

Chronic diseases are now the leading cause of sickness, disability, and death worldwide — attributable to 68% of all deaths and 43% of the global burden of disease in 2014 [535]. By 2020, these figures are expected to rise to 73% of all deaths and 60% of the global burden of disease [531]. This is the case in both “developed” and “developing” countries. For instance, chronic medical conditions affect 46% of the United States population [13], and 82% of deaths attributed to chronic diseases occur in low- and middle-income countries [535]. Apart from mortality, most chronic diseases also negatively impact a person’s functional abilities, productivity, and overall quality of life [322].

These statistics about prevalence are not only distressing in and of themselves, with respect to the value of human life, but they also foreshadow an unsustainable financial burden [534]. Already, approximately three-quarters of all health care expenditure in the U.S. is on patients with one or more chronic conditions, many of which are preventable [82]; and over the next twenty years, such conditions are expected to cost more than $47 trillion globally [52] as their prevalence continues to increase worldwide [439].

In the face of these sobering personal and societal costs, the health domain is primed for a shift toward a more proactive and personalized model of care [96], where innovative services empower individuals to better manage their lifestyle
choices, behaviors, and overall wellness [439]. Personal technology has the potential to support this revitalization of health care solutions. Sensor, web, and mobile technologies in particular could fundamentally change how we monitor and try to positively influence behavior. A powerful combination of factors is enabling this change.

First, personal technologies are becoming increasingly sophisticated, in terms of both their data-capture features and interactive affordances. Sensors now standard on most smartphones include the accelerometer, compass, GPS, gyroscope, ambient light detector, proximity detector, dual microphones, and dual cameras [274]. A variety of other personal devices are now arriving on the scene as well (e.g., wearables), and their sensing capabilities are progressing similarly. Such functionality permits broad-scale, naturalistic collection of personal health-relevant data in an extremely granular, unobtrusive, and affordable way. The technical ability to observe behavior continuously and in context also makes it possible to tailor health interventions to optimize their effectiveness for an individual user; plus these technologies provide an interactive medium through which health applications can deliver that information.

Second, recent years have seen a massive swell in personal technology penetration, which is becoming increasingly accessible and affordable. The adoption rates of mobile phones are particularly striking. In the United States, over 90% of people own cell phones and 72% own smartphones [401], and 80% of adults globally are estimated to have a smartphone by 2020 [153]. Adoption is even higher for young generations — 86% of 18–34 year olds in the U.S. own a smartphone and are heavily habituated users [64]. Smartphone ownership does decline with age, but that trend is changing over time; and other tradition-
ally underserved groups in the U.S., including racial minorities or those with relatively low income and education levels, rely heavily on their smartphones for numerous important life activities [461]. Thus while the arrival of the internet was accompanied by a “digital divide” that would have hindered the reach of health interventions to some individuals, mobile phones have been widely adopted across socioeconomic and demographic groups. Mobile network coverage is continuing to expand as well. Over 85% of the world’s population is now covered by a commercial wireless signal (surpassing many other forms of infrastructure such as paved roads, electricity, and wired Internet), and over 70% of today’s wireless subscribers live in low- and middle-income countries [244]. Such penetration is highly encouraging, especially given that it provides avenues to engage with traditionally difficult to reach populations [233].

Finally and importantly, not only has the rapid development and spread of technology placed a powerful health management platform in many pockets, but people are also receptive to using these systems for self-monitoring [431]. Individuals are increasingly using technology to measure and record a variety of health-related items; for example, 7 in 10 U.S. adults now track at least one health indicator (e.g., blood pressure, mood, etc.) for themselves or for a loved one [168, 169]. There are now over a quarter of a million health apps available to smartphone users, and it is expected that over 1.7 billion people will have downloaded health apps by 2017 [413]. Additionally, adoption of health-oriented devices (e.g., wearables like Fitbit or Apple Watch, “smart” objects like networked scales or mattresses, etc.) is quickly climbing [299]. In fact, market research indicates that such personal health management technologies will surpass $70 billion by 2024 [201]. This increasing engagement is also reflected by the rise in movements like “lifelogging” or the “Quantified Self”, which refer
to an individual using technology for self-tracking various aspects of daily life, often with a focus on health and wellness improvement. Indeed, scholars have recognized a shift in mindset from “My health is the responsibility of my physician” toward “My health is my responsibility, and I have the tools to manage it” [475]. Further, medical professionals, governments, and international organizations (e.g., National Council for Behavioral Health, World Health Organization, United Nations) also express receptivity to technology-based treatment protocols, in many cases even strongly advocating for deeper integration of personal self-monitoring tools into existing public health care systems [244].

Altogether, this swift spread and rapid technical evolution of computing technology, together with the growing engagement from individuals who want to proactively improve their health and quality of life, creates unprecedented opportunities for HCI researchers to develop robust, interactive, and tailored health applications that are also scalable and cost-effective [416, 489]. Creating such technological solutions that empower users to monitor, manage, and improve personally-significant aspects of their wellness is thus an area of immediately compelling opportunity.

This dissertation examines promising HCI efforts already underway (e.g., behavioral intervention technology, persuasive computing, personal informatics); and in doing so, I identify a gap between evidence and practice that indicates more work is still needed to make the types of personal data collected, the health metrics modeled from that data, and the behaviors targeted for intervention more personally, contextually, and clinically relevant. The crux of this dissertation then lies in “domain-driven” research strategies I provide to close this gap.
Specifically, I set forth a domain-driven development framework: a vision and plan for connecting the creation of personal health informatics technology more tightly with domain knowledge — the concepts, theories, empirical findings, user feedback, and any other forms of pertinent background information that can inform development choices regarding data collection, health modeling, and behavioral feedback.

To demonstrate this framework in practice, I use a case study I have undertaken, aimed at advancing a class of chronobiology-driven technologies for supporting sleep, daily performance, and emotional wellness. Much of this work has been published in prominent conferences (e.g., the ACM Conference on Human Factors in Computing Systems — CHI, the ACM International Joint Conference on Pervasive and Ubiquitous Computing — UbiComp, and the International Conference on Human-Computer Interaction with Mobile Devices and Services — MobileHCI) and journals (e.g., Human-Computer Interaction, Computers in Human Behavior, Assessment, and the Journal of the American Medical Informatics Association — JAMIA) or is in print, under review, or in preparation for future submission.
1.2 Dissertation Overview & Contributions

This dissertation makes methodological, empirical, and human-computer interaction contributions, organized into the following six chapters.

Chapter 1 ("Introduction") has provided the high level motivation behind my dissertation research — an alarming incidence of chronic disease, opportunities for personal technology to help address this well-recognized health crisis, and a domain-driven strategy I propose to make health management technology more efficacious.

Chapter 2 ("Background") offers a view, from an HCI perspective, of the current landscape of technologies working to pursue those opportunities. Overviewing various areas of interrelated work, I provide definitions and a common vocabulary used in the remainder of the dissertation. In its review of extant work and prominent systems, Chapter 2 also explains in detail the value in developing new generations of tools whose design choices are more deeply grounded in domain knowledge — from the data a system senses automatically or allows a user to manually log, to the health metrics it extracts from that data, to the personally and clinically relevant feedback it provides. After constructing a working definition of “domain knowledge”, this chapter concludes by specifying the components of a domain-driven development framework, which is unpacked and demonstrated in the chapters that follow.

Chapter 3 ("Domain Inquiry") explains the initial, foundation-building steps in developing a domain-driven system. Specifically, this chapter describes a process for selecting a compelling application area, identifying salient domains from which to gather background knowledge, and using that under-
standing to inform subsequent research and development decisions. To demonstrate this process in practice, I use my research on technology for supporting sleep, cognitive performance, and emotional wellness as a case study. I first explain how I recognized these as areas ripe for technological solutions and what limitations I saw in extant related work. I then describe how I identified chronobiology as a relevant domain, and I overview the background information I gathered — knowledge I hope additionally serves as a standalone resource useful for others interested in doing chronobiology-driven work. Finally, I apply this knowledge in order to plan the subsequent PHI work, particularly regarding the scope and modeling strategies for health assessment (i.e., what to assess, for whom, and how).

Chapter 4 (“Domain-Driven Health Assessment”) contributes the methods and findings from two experiments I conducted to execute the domain-informed analytic plan devised in the previous chapter. Specifically, the first experiment explored how social sensor data can be used to detect sleep-related behaviors and circadian disruptions, and it took preliminary steps toward analyzing the impact of inadequate sleep on cognition and mood. Digging deeper into daily functioning, the second experiment then built on chronobiology about cognitive performance rhythms in order to explore and interpret a number of relationships among smartphone app use, alertness, sleep, and latent biological traits. This chapter reports on the outputs of this research, along with its implications for the general framework.

Chapter 5 (“Domain-Aware Intervention Design”) presents a cycle of preparing, creating, and evaluating designs to deliver interventions and other user-facing feedback about personal health. To demonstrate this process in
the context of this dissertation’s main case study, I apply knowledge gathered during domain inquiry, findings from the two experiments presented in the previous chapter, and interactions with users in order to devise a series of chronobiology-aware design guidelines, mockups, prototypes, and full systems for supporting sleep, activity scheduling, and mental health management.

Chapter 6 ("General Discussion and Conclusion") concludes this dissertation by synthesizing takeaways and pointing out possible future directions. In particular, this chapter discusses strategies for balancing key tradeoffs, which I encountered in my case study research but are applicable to PHI more broadly: technology’s impacts on health as a double-edged sword, integrating manual and passive health management activities, avoiding over-personalization when tailoring experiences, and combining domain-driven with data-driven approaches. In closing, I describe opportunities for future work, including moving beyond the traditional single-user model to more collective styles of informatics as well as identifying areas beyond health where my domain-driven framework could generalize.
CHAPTER 2

BACKGROUND

This chapter reviews and critiques academic and industrial work on technology aimed to support personal health. Specifically, after describing how modern medicine evolved to become what it is today and the role personal technology has played, I coin a new term with a tractable definition that I use in this dissertation to refer to a class of systems that support personal management of healthy behavior. I also outline requisite features of this technology and provide examples of extant systems that possess these characteristics.

In doing so, I also motivate a core argument of this dissertation: that such technologies will have more successful health outcomes if they are made more personally, contextually, and clinically relevant — something that can be achieved by using domain knowledge to inform development choices, from the types of personal data collected, to the health metrics modeled from that data, to the designs of delivered feedback and interventions.

My intention in the following sections is to provide a context for my proposed domain-driven framework, rather than an exhaustive review. When available, I point to other sources that provide more encyclopedic reviews of the literature.

2.1 The Roles of Technology in Supporting Health

As explained in the previous chapter, technology provides a tremendous opportunity to address modern health care challenges, combat chronic disease,
and overall improve health and wellness on a broad scale. This opportunity has been recognized across numerous fields with ties to health, so such efforts unsurprisingly go by various names. In this section, I overview terminology, definitions, and relationships among these related, sometimes interchangeable, sometimes overlapping, sometimes distinct labels. For the sakes of scope and relevance, I focus on concepts and technologies most prominent within HCI and relevant to this dissertation (e.g., mobile health, personal informatics, quantified self, and so on). It is actually difficult to find a comprehensive review in the literature that provides such an overview along with a history of how these concepts originated and came to mean what they do today. I therefore find it useful to look at these ideas through a historical lens, in terms of health care’s evolution over the past three centuries.

### 2.1.1 The Roots of Modern Medicine

Modern medicine has its roots in the 19th century [76], which saw changes in the conceptualization of disease and illness (e.g., from the discovery of bacteria), an introduction and increased use of statistical methods (e.g., correlational analysis and hypothesis testing), improved care practices (e.g., anesthesia, better hygiene), and the emergence of new fields (e.g., psychiatry) for previously undertreated conditions such as mental illness [126, 154].

In the 20th century, medicine became increasingly professionalized and saw new formal mechanisms for regulating medical practice. Organized medical research also gained tremendous momentum. In the decades just prior to the start of the 20th century, the precursor to today’s National Institutes of Health
(NIH) was established, and government initiatives began to support the funding and supervision of scientific medical research [363]. Such professionalization also led to breakthroughs in public health research and programs. As public health became an increasing priority during the past century, population-level epidemiologic studies became more feasible due to instituted changes such as the deployment of periodic standardized health surveys (e.g., the U.S. National Health Survey, established in 1956) [474]. Studies also moved beyond only measuring disease prevalence to more controlled experimentation (e.g., randomized clinical trials), beginning in 1948 with the first modern clinical trial of an antibiotic drug for tuberculosis [498]. Indeed, pharmacology in the 20th century became increasingly sophisticated, with more focus on the development and use of drugs and medications (e.g., antibiotics, vaccines, psychiatric drugs, and vitamins) [474].

Numerous advancements in medical technology were achieved as well, for instance beginning with the dissemination of the X-ray machine to most hospitals near the beginning of the 20th century [496]. The technologies that are most relevant to this dissertation also have their foundations in the 20th century, during which time technological development continued to rapidly progress.

Enabled by the production of 20th century telecommunication technologies, telehealth (also known as virtual health care) refers to the use of telecommunication services and information technology to deliver health care or health information from a distance. Though telehealth is often considered synonymous and used interchangeably with telemedicine, they can be distinguished, as telemedicine refers specifically to traditional clinical diagnosis and monitoring delivered remotely (e.g., supporting a physician’s remote monitoring of a
patient or enabling transmission of medical images for diagnosis), while tele-
health has a broader scope that includes the distribution of a range of non-
clinical health services in addition to or aside from clinical services.

Telehealth is further encompassed by eHealth [378], a recently defined um-
rella term that describes the use of electronic technology or digital data to sup-
port health care, whether it be used remotely or locally [125]. Beyond telehealth
services, other eHealth technologies include electronic health records, clinical
decision support systems, and computerized physician instruction tools. The
use of such information technologies in health care is often referred to as health
informatics (also known as health care informatics or medical informatics),
which is concerned with the collection, storage, retrieval, management, and use
of health information by a patient’s care providers.

eHealth also includes mHealth applications. mHealth (for mobile health)
broadly refers to the use of mobile phones or other wireless devices to sup-
port health care [244]. A sizable portion of my research would be considered
mHealth, as I too leverage mobile technology (particularly, smartphones) to
assess health behaviors and deliver behavioral interventions. I tend to focus
on participants from the United States due to access, and mHealth is certainly
applicable and can have significant impacts in industrialized nations and high
income areas. However, given the substantial and widespread penetration of
mobile phones, a recent thrust of the mHealth field has focused on applica-
tions in developing, rural, or low-income areas, with the goal of improving
health care quality and access for underserved populations. Examples of typical
mHealth technologies include phones, mobile applications, personal digital as-
sistants (PDAs), and patient monitoring devices (e.g., network-enabled glucose
monitors or portable electrocardiogram devices); and recent years continue to see the arrival of various “smart” mHealth technologies such as smartwatches, smart eyewear, and smart scales.

Such technological developments and new treatments have helped global life expectancy rise by over thirty years in the past century [531]. However, as the prevalence of illnesses common in the early 20th century (e.g., measles, rickets, typhoid fever) shrank, an incidence in chronic diseases (e.g., diabetes, heart disease, cancer, and depression) increased [439, 535]. A greater focus in the later half of the 20th century on treating such non-communicable diseases paralleled the emergence of behavioral medicine, an interdisciplinary field concerned with developing behavioral and biomedical knowledge and techniques relevant to health in order to support illness diagnosis, treatment, and prevention [252, 450].

2.1.2 The Age of Behavior Change

I therefore designate the 21st century as the age of behavior change. One of our greatest present health challenges is dealing with mental health problems, chronic conditions, and lifestyle diseases, which are linked with how people live their lives and are the leading cause of sickness, disability, and death worldwide — attributable to 68% of all deaths (and 82% of deaths in low- and middle-income countries) [535], as noted in the previous chapter. Similarly, the top risk factors for premature death (e.g., high body mass index, physical inactivity, unhealthy food choices, smoking, and excessive alcohol consumption) all relate to lifestyle choices. According to the U.S. Centers for Disease Control and Pre-
vention (CDC), eliminating just three risk factors — inactivity, poor diet, and smoking — could prevent 80% of heart disease cases, 80% of type 2 diabetes cases, and 40% of cancer cases [324].

Consequently, scholars are reaching a consensus that “the single greatest opportunity to improve health and reduce premature deaths lies in personal behavior” [449]. Indeed, while genetics, behavioral patterns, socioeconomic circumstances, environmental exposures, and the quality and efficacy of health care all influence our health [439], research now shows that an individual’s behavioral choices (40%) combined with social (15%) and economic environments (5%) contribute the most (altogether, 60%) to personal health status [449]. Addressing such behaviorally rooted issues therefore requires getting individuals more directly involved in their own care.

Thus, in light of the rise in chronic conditions and the major role that lifestyle choices play in condition development, progression, and outcome, two major shifts are occurring within the health domain. One, it is increasingly moving toward care models focused on management rather than treatment, with monitoring as a daily activity rather than an occasional consultation. In addition, approaches are increasingly focusing on individually-targeted behavioral interventions. Technology is therefore seen as a highly appealing mechanism for delivering such interventions, as it can provide behavioral coaching that is continuously available and delivered directly to the patient. An additional benefit of contemporary technology is that it can reach populations normally unable to access care due to various financial and physical barriers [337, 338].

Based on established behavior change counseling principles known as the “5As” (assess, advise, agree, assist, arrange) [182, 522], interactive behavior
change technologies (IBCTs) were one of the first technology-based approaches to using hardware and software (e.g., DVDs, PDAs, emails, phone calls, and patient-centered websites) to deliver the 5As to patients before, during, and after primary care visits [181]. For instance, IBCTs have been used to improve diabetes self-care by delivering walking interventions, encouraging medication adherence, and facilitating patient-to-patient peer support [398].

Over time, additional terminology has been introduced to refer to the use of technology to support behavior change, and novel digital mediums continue to be appropriated for the task. Today, the term behavioral intervention technology (BIT) is broadly used to refer to a range of modern modalities (e.g., mobile phones, web 2.0, and wearable and environmental sensors) utilized to support users in changing behaviors related to physical health, mental health, and overall wellbeing [337, 339]. Persuasive technology (or persuasive computing) is similarly intended to change or maintain attitudes and behaviors [164]; and while persuasive technologies are found in many domains (e.g., sustainability, education, and activism), many today focus on applications to health, according to previous reviews as well as my own examination of academic repositories and consumer application markets [356]. Within HCI, the term “persuasive technology” is still common, but it is considered somewhat controversial due to concerns about implied coercion [406]. Researchers therefore instead sometimes simply use the phrase behavior change technologies [206].

As the landscape of these technologies for supporting healthy behavior continues to evolve, one notable characteristic is that usage is not only moving outside of clinical settings where health care has been traditionally delivered — but clinical oversight itself is diminishing. This is in large part due to the increased
availability and uptake of direct-to-consumer products for self-managing health (e.g., physical activity trackers, internet-connected scales, and bluetooth heart-rate monitors), which typically provide mobile applications or online counterparts to provide guidance in adopting or maintaining healthy behaviors [99].

A primary component of these self-monitoring systems is the collection of personal data, through self-tracking done either manually or automatically via sensors (e.g., accelerometer-based step count sensing). The term personal informatics (PI) was coined within the last decade to refer to technologies aimed at helping users collect and reflect on personal information [284]. This work conceptualized self-tracking as a five stage iterative process through which a person will prepare what data to collect and how, collect that data, integrate and organize collected data, reflect upon and interpret data, and determine how to convert gained understanding into a plan for action. Subsequent work has expanded this model in a number of ways, including to identify additional styles of tracking (e.g., goal-driven and documentation-based activities [431]), stages of tracking (e.g., a maintenance phase and lapsed tracking [149]), and ways to accommodate clinician-patient collaborations when self-tracked data from commercial tools is used as part of treatment [97].

While people use personal informatics tools for various reasons (e.g., to track finances, document visited locations, out of curiosity, or to receive rewards on social networking sites) [149], a majority of users are interested in capturing and accessing health data. Similarly, while improving self-knowledge is nominally the goal of PI, the value in exploring personal data stems for most individuals from a desire to translate gleaned information into self-improvement strategies, particularly with respect to health behavior change [476].
A recent rise in the practice of self-tracking is known as the **Quantified Self (QS)** movement, which refers to self-monitoring any aspect of one’s life by capturing physiological, behavioral, emotional, cognitive, or environmental information, typically with the goal of improving or optimizing physical or mental health. Compared to personal informatics users in general, the QS community has been identified as an example of “extreme” self-trackers [94], aiming to make the body a more knowable and hence “calculable and administrable object” through QS activities [476]. Indeed, these “QSers” capture a vast range of personal data, sometimes using highly invasive methods (e.g., brainwaves via EEG, neurotransmitter measurement via urine and blood serum, video records of one’s totality of experiences using *lifelogging* [135] apparatus, or, for one super-self-tracker, even the intensity of every crying session experienced over a period of more than one and a half years [519]).

Such extensive self-monitoring pertains to a person’s interest in conducting a form of *self-experimentation* [94]. This practice has a long history in medicine and psychology, where doctors have traditionally volunteered for ethical reasons as the first subject in a human experiment that has unknown risks [10]. Very recently, HCI researchers have begun designing technology to support what they similarly refer to as self-experimentation in self-tracking, for instance to assist an individual with Irritable Bowel Syndrome in identifying foods that trigger symptoms or to help a person determine whether timing her exercise for the morning does result in more energy later in the day [241]. This work is motivated by the fact that people want to use personal informatics tools to answer specific questions like these about their health, but current tools typically fail to effectively support such *diagnostic* self-tracking [240]. For example, many tools output graphs of raw data that users find difficult to interpret or act on [148],

---

18
and tools generally do not support personal experiments that have sufficient methodological rigor [94].

Self-experimentation technologies essentially help a user self-administer a controlled experiment; the tool creates an experiment scheduling, encourages adherence to experimental conditions, and automatically runs statistical tests from which a user can draw causal conclusions. The experiment follows a single-subject design (also known as an \textit{n-of-1 study}), which is sensitive to individual differences and where a person serves as his or her own control [286].

HCI’s interest in n-of-1 style personal informatics coincides with calls from within the medical community to adopt models of \textit{personalized medicine} (or \textit{individualized} or \textit{precision medicine}) that focus on individual, rather than average, responses to particular treatments. Such an approach can be advantageous compared to methodologies involving larger samples (e.g., randomized controlled trials), which can lead to therapeutic solutions that are beneficial to some patients but minimally effective or even detrimental for others [176]. For example, some routinely used medications benefit as few as 1 in 50 individuals, while other drugs have been found to be harmful for entire ethnic groups since clinical trials are often biased toward white Western participants [447].

\section*{2.1.3 The Individual as the Nexus of Health Management}

In recent decades, modern medicine has thus been marked by three key developments that are highly relevant to this dissertation and that, in a nutshell, have drawn the health domain toward a more individual-centric paradigm focused on wellness and prevention.
First, health care is increasingly focusing on behavior change, given the aforementioned growing crisis of chronic diseases, which are often preventable and manageable through lifestyle changes. At first, efforts to promote healthy behavior largely focused on public awareness campaigns to increase health literacy or policy-making to decrease unhealthy behavior (e.g., smoking, package labeling); but such generic methods are only modestly effective [439].

Second, recognizing the inadequacy of such one-size-fits-all approaches to behavior change, health care has become more personalized. Over the years, health care has become more patient-centered [508] to resemble more of a partnership between practitioners and patients aimed at ensuring patient preferences are incorporated into clinical decision-making and that patients have the needed support to participate in their own care [219]. Similarly, personalized medicine now more explicitly considers such individual differences in patient needs in order to account for the fact that a given treatment may not affect everyone equally. While studies on groups of people have dominated medical science over the last century, with the randomized controlled trial considered a gold standard for evaluating an intervention’s efficacy, we are now seeing increasing momentum behind n-of-1 study designs, where the sample is a single participant: the patient requiring treatment.

Finally, technological developments have allowed health care to move outside of clinical settings and have supported behavior change that is self-driven. While technology was quickly recruited as a way to more broadly and affordably deliver health care directly to an individual, physicians were still the predominant point of care until very recently. Today, thanks in part to the advent of novel devices and mediums (e.g., smartphones, wearables, and web 2.0 tech-
nologies), individuals can independently collect, analyze, and act upon data representing health and wellness. Such activities can complement — or even replace — interactions with health care professionals [433], who individuals no longer need to depend upon to introduce the notion of behavior change nor drive progress. Less reliance on doctors could additionally help buffer the negative impacts of anticipated physician shortages. Though, it is worth noting that for some conditions, professional oversight may be an important aspect of care; in such cases, the goal of technology would be to supplement and extend physicians’ efforts rather than to bring them out of the loop entirely.

In exploring how to provide each person with the tools needed to support everyday health-related behaviors, this dissertation champions this idea of personalized self-management, where care centers on the individual, who is the focus of empowerment and the nexus of positive change.

2.2 Personal Health Informatics (PHI)

As just reviewed, a variety of terminology is used to refer to various related types of technologies designed to support behavior change (e.g., “behavioral intervention technology”, “persuasive technology”, “personal informatics”, and so on). Some terms are used interchangeably, some terms are considered subsets of other terms, and some terms share a portion but not all of their characteristics with other terms — plus some terms do not have definitions that are even unanimously agreed upon.
For the sake of consistency and mutual understanding with the reader, I therefore find it useful to introduce a new term I will use throughout this dissertation to refer to the technology my research is aimed at developing, Personal Health Informatics, defined as follows.

Personal Health Informatics (PHI) refers to technology that supports personal management of healthy behavior. In essence, this class of tools (1) facilitates collection of personal data, (2) enables analysis of information to assess a targeted aspect of health, and (3) provides feedback to help a person gain self-knowledge and potentially change or maintain behavior accordingly. While this definition could be applied to a range of application areas, this dissertation is particularly focused on the development of PHI systems that promote healthy sleep, cognitive performance, and emotional wellness.

First, I include the word “informatics” in the term, as a primary goal is aiding a user in capturing, interpreting, and acting upon personal data. Since this dissertation is focused on individuals engaged in this process for reasons related to “health”, I include that word in the term in order to distinguish PHI from the broader definition of “personal informatics”, which as explained earlier, can encompass a wider range of personal data from finances to energy consumption. Finally, I include the word “personal” because (as described at the end of the previous subsection) the focus is on personal usage that places an individual at the center of services that enhance health by promoting self-management.

Because my work comes from an HCI foundation, the characteristics of PHI are inspired by and most closely resemble tools that grew out of the HCI field, namely personal informatics (PI) tools (compared to, say, telemedicine tools).
The key differences are that PHI is scoped to health self-management, more explicitly designates the health assessment element (whereas PI traditionally focuses on users’ data collection and interpretation practices and less on what is potentially happening under the hood), and aims to provide user-facing feedback that is more personalized based on those user models (whereas PI tends to provide generic visualizations [21]). At the end of the day, a main reason for introducing a new term in this dissertation is simply to have a tractable definition with a clear-cut set of characteristics, given that prior definitions are either broader than necessary for this dissertation’s purposes or somewhat discrepant depending on the source doing the defining.

Specifically, I consider (1), (2) and (3) as the fundamental components to “qualify” a system as PHI:

1. **Capturing Personal Data**: By harnessing emerging mobile and ubiquitous technologies (e.g., smartphones, sensors, Internet of Things, pervasive network coverage, etc.), rich datasets about personal behavior can be collected in context, through manual self-report and automated sensing.

2. **Analyzing Data to Assess Health**: From this personal data, health metrics can be computed, condition symptoms detected, and predictive models of personal behavior built.

3. **Delivering User-Facing Feedback**: Given this model of an individual’s health and contributing behaviors, tailored feedback can be delivered in order to support personal self-management and, in turn, enhance overall wellness.
Thus essentially, personal data is the system input, user-facing feedback is the system output, and in the middle, analysis transforms data into information. Though, these elements can be overlapping and iterative; that is, there are not necessarily hard boundaries between them. For example, many passive sensing systems intertwine data collection and data analytics through a continuous loop in which user models retrain themselves based on fresh data that is sensed in real-time. Or, a system can collect a user’s positive or negative reaction to a piece of feedback — whether it be implicit (e.g., acting upon a behavioral recommendation) or explicit (e.g., using an option to “like” a recommendation) [247] — to help profile that person’s preferences. Through such an iterative process that incorporates new data, refines models, and updates feedback, a system can continue providing support appropriate to a user’s evolving needs over time.

Further, I find it useful to think about each of these aspects — capture, analysis, and feedback — in terms of the level of responsibility and agency placed in the hands of the user versus the system. Other researchers have used labels such as how “participatory” [453] or “cooperative” [379] a system is, in characterizing this relationship between its autonomy compared to that of the user. First, personal information can be collected manually, using automated approaches, or through a hybrid combination that augments self-tracking with passively sensed data. Then, analyzing that data can be left up to the individual, who is responsible for exporting, organizing, and making sense of that information; or, a system might help integrate data, compute health statistics, or apply machine learning algorithms to pull out patterns that the user can then examine. Finally, feedback may be presented in more descriptive ways that are open to interpretation by the user; or, feedback may be more prescriptive in dispensing explicit directives.
In the following subsections, I describe each of these components (capture, analysis, feedback) in more detail. Throughout, I include examples of related systems in order to demonstrate how these qualities can look in practice; I focus on work from the HCI literature, as that is the community to which this dissertation primarily speaks.

### 2.2.1 Capturing Personal Data

To begin, any personal health informatics tool requires input data. This input provides details about the user’s behaviors, environment, or any other personal attributes relevant to the health outcome the tool is designed to support. This data can be captured manually by a user, automatically by sensors, or through some hybrid approach. This subsection overviews these ways PHI technology captures data, providing examples and pointing out advantages or drawbacks.

#### Manual Data Collection

People have manually self-tracked health-related personal information long before digital tools existed to explicitly support the activity. In the 1940s, clinical research began using written diaries, in which people self-report symptoms and health actions as they occur [9, 501]. While such pen-and-paper approaches are familiar and easy to use for many people, they also face well-known limitations including the risk of forgetfulness, retrospection errors, uncertain adherence, and lack of response-time information [55]. Numerous researchers have worked toward addressing these limitations, and over the past thirty years or so, much attention has been specifically focused on how technology can help.
At first, studies used digital devices such as pagers [118], pre-programmed wristwatches [291], or text messages [15] to deliver reminders to record information, though the recording itself was still made on paper. This sort of prompted self-report has traditionally been associated with experience sampling (ESM) [114] and ecological momentary assessment (EMA) [470], which are methods used to collect information about various aspects of daily life in the moments they are being experienced. Such in-situ assessment can reduce retrospective biases, help in identifying ecological factors that may be contributing to an individual’s health status, and reveal whether a person is employing acquired health management skills (e.g., from clinical therapy) in everyday life [420].

More recently, technology advanced to a point where it could be used not only to prompt users to record information but as the recording medium too, instead of paper. Early electronic data collection methods included handheld computers such as PDAs loaded with questionnaire programs [33, 452]. Today, most HCI researchers working on improving the experience of manually tracking with technology focus on smartphone-based self-report and target activities that are quite difficult to detect automatically such as pain levels [491], food intake [44], or subjective wellbeing [318].

One thread of HCI work is particularly interested in developing more “lightweight” self-tracking interfaces. Several allow information to be recorded directly from a smartphone’s lockscreen so that a user does not have to launch a full diary application or even unlock the phone. For example, the SleepTight smartphone app allows a user to tap icons on the lock screen in order to journal sleep disruptors [93]. The Slide-to-QuantifySelf app also allows a user to record health metrics (e.g., daily water intake) using lock screen widgets [490],
and the LogIn system similarly repurposed unlock interaction gestures to capture self-reports for sleepiness and mood [548]. In developing the MoodRhythm smartphone app for managing bipolar disorder, colleagues and I used notifications to prompt momentary self-reports, and we made it possible for patients to quickly record target variables directly from the notification panel [312].

The manual capture of data is associated with several benefits. Self-tracking can empower users with a sense of agency [354], plus directly engaging with data can foster self-awareness [45, 94]. The “obtrusiveness” of the practice is precisely its main advantage, as this is what enhances mindfulness about behavioral choices and can promote adherence to behavior change goals [260, 262].

However, manual self-tracking is associated with disadvantages as well. Foremost, self-report can be burdensome [104] due to the time and effort it requires. This is a particular challenge if a technology is intended for long-term use (e.g., to manage a chronic health condition). By reducing the amount of information a user needs to report, briefer assessments can overcome this burden somewhat; but the shortness of such instruments can weaken their validity [73]. Burden can also degrade the quality of data collected if it translates into inadherence (i.e., missing data), which can also result from basic forgetfulness. This issue is compounded if it leads to subsequent backfilling of missing data, which is common [471] and typically inaccurate due to retrospective recall biases [112] or even fabrication [55].

Data inaccuracy can also occur in cases where a person’s capacity for reliable self-assessment is compromised, such as when suffering from sleep deprivation [139] or if experiencing symptoms of certain mental health conditions including bipolar disorder [190]. Relatedly, the act of self-assessment can be impeded in
certain contexts. For instance, self-report is simply not possible or could be dangerous in some situations (e.g., during sleep or while driving); or performing self-tracking may be difficult in some social settings, especially if the individual is reporting about a potentially stigmatic health condition.

Next, while increased self-awareness can induce positive reactivity that leads to desirable behavioral changes, negative reactivity can also result. For example, in the context of self-reporting pain, some evidence finds that recalling pain and coping strategies can lead to positive outcomes such as an increased sense of control over the pain [204]. However, other findings suggest that repeated self-assessment of a potentially traumatic situation can actually draw one’s attention to and foreground negative perceptions, thereby worsening the lived experience of that condition [258].

Finally, it can be infeasible for a person to capture the array and granularity of data necessary for a system to produce a sufficiently comprehensive profile about that individual, comprised of the multiple personal variables, behavioral determinants, and other indicators needed to accurately model a health outcome of interest [45]. Thus, interest arises in more system-driven approaches to data collection that are either fully automated or that complement self-report with passively captured information.

**Automated Data Collection**

With automated or “passive” data collection, physiological or behavioral trace data is captured using sensors or from technology usage logs. These sensors can be worn on the body, located in the environment, or embedded within a personal device.
Systems designed to encourage physical activity have used a variety of body-based sensors such as pedometers [106, 290], biometric sensors like electrocardiogram (ECG) [124], or custom sensing setups [107] composed of sensors to capture sound, temperature, light, and humidity among other inputs [95]. Many of today’s commercial wearable devices designed to support healthy behaviors (e.g., Fitbit, Jawbone UP, Microsoft Band) are essentially accelerometer-based wristbands that passively monitor activity and sleep [410], though newer models incorporate additional sensors, for instance to measure heart rate or galvanic skin response. Such sensors have also been incorporated into clothes and jewelry (e.g., to assess mood [494] or support remote patient monitoring [414]).

The main disadvantages associated with body-based passive sensing are the discomfort of wearing the sensing device, the limited battery life, and the fact that smaller form factors inherently constrain the sensors that can be hosted — although battery and miniaturization advances are helping to address some of these issues [410]. As with manual data collection, forgetfulness can be an issue for passive strategies as well; for instance, a user may forget to wear the sensing device.

As environment-based sensors do not face these challenges, researchers have been exploring how instrumented homes can automatically capture health data. One system automatically captures weight using a scale built into the toilet, heart rate data using an ECG monitor in the tub, and body temperature from a bed sensor [376, 479]. Others have placed sensors to automatically collect health metrics into furniture like chairs [195] or mattresses [257], onto home appliances, or into cars [261]. The notion of an “internet of things” connected smarthome would further extend such capability and connectivity to numerous
other objects within one’s living space, though this vision still remains largely unrealized.

Smartphone-based sensing, on the other hand, has not only garnered tremendous interest recently from researchers in HCI but has made substantial progress, with mobile sensing emerging as a field in its own right [274]. The mobile phone has rapidly evolved into a powerful computing platform, with a variety of sensors now standard in consumer smartphones for automatically capturing motion (e.g., accelerometers, gravity sensors, gyroscopes); location (e.g., GPS, orientation sensors, magnetometers); and environmental data (e.g., barometers, photometers, thermometers, cameras, microphones). Mobile sensing is often applied to the area of health, with recent reviews overviewing prominent health-oriented smartphone sensing systems [86, 255]. As part of this dissertation’s main case study is focused on sleep, I provide a more comprehensive review of sleep sensing systems in particular in Section 3.2.3.

A novel twist on sensing is “soft sensing”, which passively captures data not from hardware based sensors but from software usage logs. That is, as a byproduct of interacting with numerous technologies (e.g., computer programs, smartphone apps, or web services), individuals generate a massive amount of data (often referred to as “big data” — or, “small data” if referring to the data generated by a single individual [152]). By leveraging such digital trails, researchers taking a soft-sensing approach have had success in inferring a user’s health-related behaviors, contextual or psychological states, or other personal characteristics of interest [120].
Overall, automated sensing helps relieve user burdens by reducing both the time and mental overhead associated with self-tracking, plus sensed data is often more accurate and granular than manually tracked data. Automatic sensing can also capture informative quantitative signals that are relevant to health but may be imperceptible to the person generating those signals [523]. However, sensors can be privacy invasive or (for wearable sensors) uncomfortable to wear, and they can reduce personal awareness about collected data [283]. In addition, automatic data collection can work well to acquire “objective” information like heart rate or location; but as mentioned, it does not lend itself as well to measurement of subjective experiences. Finally, practically speaking, while capturing some types of data is now reliable (e.g., location, walking detection), other types of data are still more elusive.

Hybrid approaches attempt to make the best of both worlds by employing more than one capture mechanism. For example, UbiFit automatically infers physical activity data about walking, running, and cycling; but the system allows the user to add activities it cannot automatically track like yoga or swimming [108]. The Somnometer sleep support system automatically tracks sleep duration from phone data but obtains sleep quality information from the user [455]. Commercial health trackers like the Fitbit and Jawbone UP similarly use accelerometers to automatically track activity and sleep data while also providing an interface where users can manually log additional information such as mood, meals, caffeine, and water intake. Such information is notoriously difficult to automatically detect; and a number of self-monitoring technologies use this hybrid, “semi-automated” approach to manually collect data that cannot be otherwise captured. In Section 6.2.2, I suggest other potential strategies for integrating manual and passive data capture methods.
2.2.2 Analyzing Data to Assess Health

Once a user’s personal data has been collected, it can be analyzed to derive relevant health metrics and user profiles. This process is sometimes referred to as data “mining”. The aforementioned model of personal informatics use would consider this process as a part of “integration”: preparing and transforming the data into information that can be reflected and acted upon, which is especially necessary if the format of collected data is different from the format required for reflection [284]. This process can again be driven more by the user or by the technology.

Manual Analysis

Manual analysis involves an individual examining his or her data directly, for instance, by sifting through daily weight measurements to pull out overall trends or to compute correlations with other tracked personal data (e.g., physical activity, calorie intake, or numbers of hours slept). Quantified Selfers are one group particularly known to engage directly with their raw data in order to analyze it “by hand”. One recent study found that nearly half of its QS participants put their data into a spreadsheet tool like Excel or Google Spreadsheet that could assist them in manually running simple statistics, and over one-third of participants even programmed custom software to run such analyses [94]. More anecdotally, on the Quantified Self community’s main blog and forum, one can find numerous posts (e.g., [407]) that convey guidelines or personal experiences about conducting manual analyses in order to identify patterns, discrepancies, or health predictors from data like step counts, dietary intake, and medication.
Automated Analysis

In more fully system-driven analysis, technology automatically processes tracked data on which it runs statistical analyses. For example, Health Mashups [45, 488] automatically computes statistical correlations within various data streams acquired through both manual tracking (e.g., of exercise, food, mood, pain) and passive sensing (e.g., of calendar information, location coordinates, weather, activity, and weight). The BeWell system [273, 275] automatically computes a wellbeing score using data about activity, sleep, and social interaction.

Another approach is to use algorithms to infer health states using sensor data. For example, Mobile Heart Health [346] applies an off-the-shelf algorithm to a user’s real-time ECG data in order to derive a measure of stress. Given a sufficient number of data points, technology can also generate a statistical model of behavior. CenceMe [332] applies machine learning to mobile phone sensor data (e.g., accelerometer, Bluetooth proximity to other devices, GPS, microphone) to classify a user’s healthiness along with other attributes. As mentioned, a PHI technology’s input, analysis, and output processes are often overlapping and iterative; these examples demonstrate passive sensing systems where continuous data capture goes hand in hand with iterative analysis.

Somewhere in the middle of this manual–automated analysis spectrum, some tools assist with converting raw data into a more human-inspectable format. For instance, a number of systems use machine learning or data mining algorithms to do data aggregation or labeling (e.g., converting accelerometer data into physical activity categories [108]). We could say a person analyzing these pre-processed data would be doing so through a hybrid, system-supported manner rather than through a fully manual or fully automated approach.
Essentially, the process of analysis serves to translate a repository of personal data into metrics, statistics, and user models about an individual’s health. By communicating such information through a user interface, a PHI system can next support self-reflection, behavior change, and overall health management.

2.2.3 Delivering User-Facing Feedback

In addition to collecting and analyzing data, personal health informatics systems also aim to represent this information through legible feedback that connects to a person’s real-life experiences and opportunities for behavior change. This feedback can also be referred to as a health “intervention”.

Based on a literature review and my own design experiences, I have charted this design space, identifying the following primary design dimensions of feedback supplied to a user: format, delivery medium, attentional demand, room for interpretation, and level of personalization. Sometimes, dimensions can share common borders or certain aspects can even overlap, often because some design choices tend to go hand in hand with each other. For example, audio is the feedback format of a device that uses chime sounds to communicate step count information, while the volume of that audio modulates the feedback’s attentional demand. Further, this is not meant to be an exhaustive set of all the possible attributes feedback can have — for instance, it can also be important to consider a piece of feedback’s audience (e.g., private vs. public viewability), scope of input (e.g., personal, family, or community level data), permanence (e.g., temporary vs. archival), or explorability (e.g., static images vs. an interactive interface), among a variety of other possible dimensions.
Still, I believe format, delivery medium, attentional demand, room for interpretation, and level of personalization represent the key design levers to be configured when deciding how information will be conveyed by a PHI tool. Accordingly, this multi-dimensional design space can be used for prescriptive purposes (e.g., when developing a new PHI technology, in order to identify design decisions to make along with trade-offs among those choices) or for more descriptive purposes (e.g., to map an existing PHI technology onto the design space in order to characterize its feedback techniques).

In this subsection, I describe each of these dimensions in more detail, including pros and cons to different feedback strategies. In doing so, I provide thoughts around when a designer might or might not want to pull certain levers, depending on the goals of the system, the types of information it wants to communicate, and the kinds of interactions it wants to support. I also map existing PHI tools onto these dimensions both as a way to review how extant technologies typically present feedback and also in order to illustrate examples of how a system can actually embody a design dimension in practice.

Format

Feedback can be presented via any of the human senses: sight, hearing, touch, smell, and taste. PHI systems often display information visually, for instance using printed text; colored light; or charts, maps, or other visualizations. For example, the aforementioned Health Mashups system presents the results of its statistical analyses using charts (e.g., bar, line) as well as feeds of observed correlations expressed in natural language sentences (e.g., “You are happier on days when you sleep more”) [45]. MyBehavior uses eating behav-
iors and physical activity data to generate healthy lifestyle suggestions that it provides through lists of text-based recommendations about meals to have or avoid and physical activity to perform (e.g., “Small walks each hour near Garden Ave”) [408]. PHI systems running on devices with small screens can limit the amount of visual information displayed; for instance, the Fitbit displays step counts in a single number or uses LEDs that illuminate when activity or sleep goals are met, while the companion app and website provide more detailed visual feedback in the form of charts and graphs. MoodLight similarly uses light and color, (specifically, lamps with color-changing bulbs) to reflect an individual’s mood based on biosensor data [463].

Other systems use visual metaphors to communicate about aspects of a user’s behaviors. For example, Gluballoon, a diabetes monitoring application that runs on a wearable display, uses the metaphor of an animated hot air balloon to illustrate a patient’s blood-glucose levels [137]. UbiFit’s phone wallpaper displays a garden, where the number of flowers represents a person’s amount of activities and butterflies represent attained goals [108]. BeWell similarly uses a phone wallpaper (an animated wallpaper in this case, as the newer BeWell system runs on smartphones), where the amount of fish, turtles, and fish schools reflects the user’s wellbeing scores for activity, sleep, and social interaction, respectively [273]. Along the same lines, Fish’n’Steps uses a kiosk display to reflect a person’s physical activity levels with fish characters, where the size and facial expressions of these fish map to goal progress and achievement [290]. Systems using metaphors are based on the premise that such representations are more intuitive to understand and will therefore facilitate information being more easily digested, interpreted, and acted upon.
PHI systems can also use audio as a format to communicate information. For instance, the InShape footworn accelerometer system is designed to encourage physical activity by playing pleasant chime sounds to encourage brisker walking [214]. Health apps often play a notification sound when a piece of feedback is available or sometimes ring an alarm to remind users when it is time to log data. Such notifications are often accompanied by tactile feedback as well (e.g., vibrations). More generally, such “haptic” formats use the sense of touch to convey information. Finally, while the use of smell or taste to communicate feedback is relatively uncommon, HCI researchers are beginning to investigate such ideas. I can imagine a PHI system designed to promote healthy eating might someday change the taste of a smart spoon in order to discourage a user from eating a food it deems unhealthy.

When considering which format a piece of feedback should take, a number of factors can be considered. First, it is important to judge whether a particular format can adequately and sensitively convey the meaning held by the data it is intended to represent. For instance, visualizations like charts and graphs grew out of a scientific tradition, with these diagrams’ geometry often designed to emphasize smooth trends. However, in my work on bipolar disorder, I found that individuals managing that condition can experience a real disconnect between the erratic nature they associate with their mood fluctuations and the smoothed patterns of traditional representations often used by PHI systems.

Additionally, skills, knowledge, values, culture, and various other personal attributes can impact how an individual perceives the format of a given piece of feedback [216]. For example, children might find the more metaphorical formats more attractive and engaging [174] (e.g., a UbiFit garden of flowers or a
Fish’n’Steps tank of cartoon sea characters), while if a tool’s intended users are adults, they may prefer more traditional charts. An individual with low vision may have substantial difficulties with feedback that uses a small font, while someone who is color blind may not get much value out of a PHI system that uses color to communicate (e.g., MyBehavior’s red or green borders to indicate whether a consumed food is healthy or MoodLight’s lamps that communicate mood through colored light). Similarly, systems considering the use of audio should consider the anticipated contexts of use, as sound might introduce user concerns about privacy, stigma, or social etiquette.

**Delivery Medium**

The delivery medium refers to the device where the feedback is provided. Mobile phones are now the most popular PHI feedback delivery medium, given their ubiquity along with their capability to both capture data as well as output feedback. Other common delivery mechanisms for PHI systems include websites, wearables, public displays, or virtual reality. Homes or other buildings may be able to deliver information in the future via a variety of walls, surfaces, or other objects in one’s living or work spaces [222].

A main consideration when it comes to selecting a device for delivering feedback is optimizing the chance of information being received, especially if that feedback is time or context sensitive. This makes mobile phones attractive due to their portability and the tendency for users to keep their phones on themselves or nearby almost constantly, even during sleep [64]. Using mobile phones also relieves a user from having to carry a separate device for health management. Depending on the type of data a PHI technology is aimed at providing
feedback about, other types of devices may be more appealing to a designer, however. For example, if feedback is based on physiological data (e.g., heart rate, breathing, stress), then it makes sense to choose a feedback device that can double as a data collection device (e.g., a wrist-worn wearable embedded with sensors for capturing those physiological data). The intended contexts of use can also help guide selection of delivery media (e.g., mobile devices would be more appropriate for on-the-go usage, while a large display in the home might be better suited for a PHI system aimed at helping parents monitor their children’s growth patterns). Finally, practical issues of usability or affordability can impact whether or not a device is well-suited to delivering feedback. For example, while smartwatches were originally touted as having many of the same benefits as smartphones when it comes to portability and easy access to information, many of these devices received criticisms for being uncomfortable to wear (e.g., too thick), hard to maintain (e.g., poor battery life), or impossible to upgrade (e.g., too expensive). If a PHI system for managing a chronic condition required frequent and long-term use of a device, then such issues would be a major barrier to adherence.

**Attentional Demand**

Next, feedback can be provided using ambient, subtle cues or in more conspicuous ways. Ambient displays often have a focus on aesthetics and aim to integrate well into the environment without being distracting, while overt feedback more directly demands that a person notices and engages with it [315]. The phone wallpapers of UbiFit and BeWell take the ambient strategy; they require low attentional demand and are appreciable *just* enough that they are noticeable but not interruptive [216]. ShutEye similarly presents a glanceable timeline of
sleep hygiene recommendations on a user’s mobile wallpaper [36].

Other systems have been designed on the same premise of providing passive cues about personal data, but information is displayed in more publicly-visible mediums in one’s environment rather than on the private screens of personal devices. Many have focused on communicating emotional data with the aim of supporting awareness of personal or collective wellbeing. For example, Mood-Light uses lamps to convey mood, as mentioned. Similarly, mood.cloud [451], an interactive data-as-art installation, is composed of a touchscreen tablet that collects mood data via the Photographic Affect Meter [400] along with a large cloud-like sculpture made of LED light strands that change color to reflect the mood of the last two dozen people who interacted with it. Other research has pursued the similar idea of embedding an LED, whose color maps to an individual’s bio-sensed emotion, into a variety of other delivery mediums such as hanging lanterns and crystal charms [434].

On the other hand, some systems are more obtrusive in their feedback delivery in order to draw concentrated attention. A number of researchers have used short message service (SMS) or text messages to deliver interventions or tailored messages with behavioral suggestions (see [161] for a review), which can be timed for delivery at points in the day when an individual is particularly vulnerable or in need of information. More recently, PHI systems have begun using smartphone push notifications to deliver alerts that contain reminders to self-report data, behavioral recommendations, or updates about accomplishments (e.g., an achieved step goal). Such technologies “push” feedback to the user. “Pull” technologies would be much less demanding of attention, as they rely on individuals proactively requesting or seeking information, for instance
by taking the initiative to visit a website containing visualizations about daily physical activity progress.

Sometimes overt alerts are triggered when specific states are detected through passive sensing. For example, when the MONARCA smartphone application detects manic symptoms in an individual with bipolar disorder, a screen of coping strategies is automatically launched [31]. Devices like the Fitbit and Jawbone UP provide idleness alarms that vibrate the device or flash lights to prompt the user to move around if stationary too long. Similarly, low levels of physical activity trigger the PersuasiveSens system to send an SMS with healthy eating or physical activity encouragements (e.g., “It is a beautiful day. Go out and do brisk walking for 30 mins”) [85].

Such real-time performance feedback can be a powerful driver of positive health behavior change [246]. The emerging area of “just-in-time” intervention design is particularly interested in delivering personally-tailored, contextually-aware, and well-timed feedback in a non-irritating way. A number of recent systems have explored just-in-time prompting to motivate behavior change (see [361] for a review). As just two examples, the EmoTree smartphone app aims to help individuals avoid bouts of emotional eating [80], while the sensor-triggered iHeal uses physiological data about stress to guide substance abuse interventions [62]. Relatedly, other work has found that a person’s current affect, stress, activity, location, and the time are important in predicting one’s cognitive, physical, and social availability to attend to a delivered intervention [440].

In these ways, feedback can be designed to demand attention less or more, depending on a system’s intentions in delivering it (e.g., enhancing peripheral awareness versus preempting destructive behaviors with an urgent message).
Room for Interpretation

In mapping out these design dimensions, I observed that a tool’s attentional demand often aligns with the degree to which that tool gives guidance for interpreting delivered feedback or leaves the interpretation up to the user. This room for interpretation could also be thought about in terms of the prescriptiveness versus open-endedness of the feedback.

On one end of the spectrum, feedback can leave little opening for interpretation. For example, MyBehavior conveys dietary feedback with messages like “Avoid large meal” written in red and with a picture of a stereotypically unhealthy food. Little interpretation is needed (or afforded), as the directive is clear: do not eat that. On the other hand, a system providing more descriptive, open-ended feedback might present a bar chart of the user’s step counts across the days of the week. Such a view would leave most of the sensemaking up to that individual. Moving a bit more prescriptive, the system could add a dotted “goal” line across a point on the step count axis (e.g., at 10,000 steps, a generally recommended daily target) or paint the bars of the days falling below that number with a color associated with poor performance (e.g., red [346]). These more explicit signals serve to guide the user toward insights the system wants to ensure she gleans from the feedback (and, in turn, the behaviors the feedback is intended to prescribe).

Other research has identified a similar dimension that it terms “control versus empowerment”, where a controlling technology would make decisions for the user or automatically control the environment, while an empowering technology would present information to the user or help a user learn how to control the environment herself [222]. For example, the aforementioned EmoTree
smartphone app might deliver an empowering, just-in-time intervention message to preempt emotional eating, while a more controlling technology in a futuristic smart home might automatically lock the refrigerator door.

From a critical anthropological perspective, empowerment is considered the more advantageous approach, as to mitigate concerns about manipulation, free will, and exploitation [359]. Designing for flexible interpretation can also help avoid users from feeling that a system is passing normative judgements about them, which can feel insensitive and be demotivating. Designs that leave more room for interpretation can also facilitate self-reflection, with open-ended feedback mediating the sensemaking experience between current self and envisioned self. Such mindful introspection is valuable, as it helps a person become more aware of her current status along with how her behavioral choices impact health — self-driven insights that can help an individual learn how to make more positive choices, even beyond the scope of one particular piece of feedback, including when technological guidance is unavailable.

On the other hand, a lack of guidance in interpreting data can sometimes be troublesome for users — confusion and uncertainty about how to act on provided feedback can lead to frustrating experiences or potentially detrimental misinterpretations. A designer can therefore make choices about this design dimension by considering characteristics of a PHI tool’s intended user or targeted aspect of health. If it is possible and beneficial for a person to learn the roots of health symptoms or engage in self-directed contemplation, it may be desirable to deliver feedback that affords more interpretation; whereas, if the capacity for self-reflection is compromised in a population of interest or if humans inherently find it difficult to make sense of a given health condition, then a more
direct, close-ended form of feedback may be preferable. Further, as with all de-
sign dimensions I have presented in this section, a PHI system might consider
adapting the interpretability of its feedback over time or depending on context;
for example, a system might encourage self-reflection and personal growth in a
novice user, yet provide more explicit suggestions to a well-seasoned individ-
ual who has already come to understand the roots of her health condition and
simply needs to receive, say, fitness-related feedback that will optimize calorie
expenditure given the parameters of her current situation.

Level of Personalization

Personalized feedback is tailored to meet individual needs and expectations,
while generic feedback is designed for a more prototypical user. Sometimes, a
generic intervention may suffice. For instance, idiosyncratic differences might
have relatively little impact on some health outcomes; and generic feedback can
at least be more effective than no feedback, as found in a few studies on mo-
tivating feedback recipients to increase physical activity [71] or reduce weight
[229]. In some cases, personalization can even be problematic, for instance when
over-personalization produces filter bubbles or propagates discriminatory pat-
terns — issues I discuss in more detail in Section 6.2.3.

In most cases, however, interventions are more successful when they’re tai-
lored to accommodate individual differences, as I explain below. Furthermore,
generic feedback can actually be harmful in some contexts, especially when
dealing with vulnerable populations (e.g., patients with serious mental illness)
[136, 161, 337]. Therefore, I generally argue for the development of PHI systems
that tend more toward the personalized end of the spectrum.
Here, I discuss why individual differences matter when designing around behavior change. In fact, a number of personally-variable factors can affect whether an individual performs positively or negatively while pursuing a behavior change goal [27, 29, 296]. Similarly, the efficacy of technology-mediated behavioral interventions can be impacted by individual differences in psychological, cognitive, demographic, or contextual attributes.

Commonly explored psychological differences typically revolve around personality since it is a well-established construct with well-validated instruments available to measure it. Studies have found that personality helps determine a person’s motivation and engagement [373, 374] as well as reaction to persuasion [235, 369], including persuasion delivered through health-related behavior change technologies [198, 237].

Beyond personality, research shows that how much an individual is inclined to engage in and enjoy cognitively complex activities — inclinations linked to one’s "persuadability" [236] — can also influence that person’s receptivity and adherence when it comes to health and lifestyle related persuasive requests [77]. In addition, willingness to use eHealth technology (specifically, health education and behavior change applications) [460] has been linked to behavioral risk factors (e.g., depression) as well as demographic characteristics (e.g., age and socioeconomic status) [482]. Studies also suggest that females are more receptive to behavior change strategies in general but that there can be gender-differences in the persuasiveness of particular strategies [381].

In addition, environmental or situational factors can also affect the efficacy of behavior change tools [163, 164] — external factors that can vary across individuals or within the same individual over time (e.g., location, nearby people,
or various other contextual factors that may relate to a targeted aspect of health [128, 443]). For example, the walkability of a person’s neighborhood can impact adherence to physical activity interventions [253]. In addition, health behaviors are largely conditioned by social context, and a network of social influences can substantially aid or undermine various efforts at personal change [74]. These environmental and social variables can also interact; for example, an exercise support group of coworkers, together with the availability of worksite fitness classes or equipment, can lead to increased physical activity [513].

Consequently, given that innate characteristics as well as extrinsic variables can influence behavior and the efficacy of interventions [147, 374], it is considered important for such feedback to be personalized [46]. Personalization has been defined as “a process that changes the functionality, interface, information content or distinctiveness of a system to increase its personal relevance to an individual” [50]. Various studies indicate that personalized messages (e.g., based on a person’s health goals, motives, perceived barriers, and readiness to change behavior) improve adherence [136] and have a greater positive impact on health behavior change than untailed or bulk messages [70, 71, 161, 437]. Further, research shows that personalizing content and experiences can reduce cognitive load, improve user satisfaction, strengthen the impact of persuasion, and promote continued use [156, 381, 492]. Given that individuals respond differently to design cues [374], personalization also helps ensure a system’s feedback has its intended effect.

Personalization can be implemented by providing customizable settings or an adaptive interface [446]. A customizable system provides affordances (e.g., settings, interactive options) for a user to control this personalization. For ex-
ample, ShutEye allows users to specify custom bedtimes, which adjusts how the system generates sleep hygiene feedback. Research associates customization with positive outcomes (e.g., increased attachment, appreciation, and satisfaction with technology [156]) as well as beneficial side-effects from the expression of individuality [51, 473]. However, there are also disadvantages of customization. Ironically, choices can be formulaic, shallow, and inadequate for satisfying individuals’ idiosyncratic needs — needs into which users do not even necessarily have good insight themselves [456]. At the same time, too many choices can be overwhelming and demotivating [223], and striking the right balance in supplying options can be challenging. An adaptive system drives personalization more automatically by tailoring its functionality to a given user’s characteristics or current contextual information, often by using machine learning methods that “learn” from the user’s data. For example, MyBehavior learns its user’s preferences in order to automatically tune recommendations about healthy meals and exercise routes.

Altogether, these findings suggest one-size-fits-all approaches to PHI design may not be the best approach, given individual differences can influence the efficacy of an intervention for a given person [234] and personalized guidance is linked to more successful behavior change outcomes [75]. Across individuals — and even within individuals across changing circumstances — an array of variables can differ. Such idiosyncrasy necessitates the development of tailored tools that consider, even embrace, such differences in order to support diverse personal needs. In turn, these personalized interfaces can minimize negative user experiences while optimizing for motivation and positive outcomes.
2.3 Domain-Driven Personal Health Informatics

So how do we go about developing a personal health informatics system, comprised of components for collecting data; analyzing that data to extract health-related variables; and providing personally meaningful, understandable, and actionable feedback?

Central to this dissertation is the idea that domain knowledge can help drive this PHI development process. In this section, I explain why. I also introduce the notion of a domain-driven development framework, which supplies concrete steps for incorporating domain knowledge into HCI practice. I describe this framework in more detail in subsequent chapters, where I also demonstrate it in action using my own research as a case study.

2.3.1 Defining Domain Knowledge

In essence, knowledge is an understanding of something. Organized bodies of knowledge originated from the human need to make sense of the world around us [285]. This understanding can be theoretical or practical in nature. That is, knowledge can be generated through use of the scientific method in order to produce theories and empirical evidence — known as “scientific theory” or “formal theory”; while experiences, trial and error, or one’s own ideas about a situation produces “situated knowledge” or “informal theory”. A domain is a subject area that holds relevance to a given problem. Thus domain knowledge is a body of understanding specific to that particular area — the part of the world investigated by a specific discipline [304].
Domain knowledge can include concepts, definitions, descriptions, empirical evidence, formulas, algorithms, theories, taxonomies, guidelines, principles, methods, and procedures [98]. In this way, domain knowledge provides a set of abstractions together with a concrete vocabulary for describing, explaining, and predicting phenomena. Domain knowledge about such information and the skills for making sense of it can be acquired through education or experience, by learning, discovering, and perceiving. Someone who has such training or expertise in a given domain is referred to as a domain expert.

In the context of PHI, domain knowledge typically refers to familiarity with disciplines beyond computer science, information science, or other fields associated with HCI (i.e., training in solving problems that are typically not the primary focus of these fields’ curricula). For example, HCI researchers developing behavior change technologies might turn to health professionals with expertise in physical or psychological health such as clinicians, biomedical researchers, nutritionists, exercise physiologists, behavioral scientists, and psychologists [11]. In the context of PHI technology, envisioned users of a system can also be considered a type of domain expert, in the sense that they possess expertise in their own lived experiences and perceived needs associated with a targeted health condition.

2.3.2 The Value of Domain-Driven PHI

To date, PHI research has focused primarily on behavioral theory as a source of domain knowledge. This is a fitting choice, as behavioral theory provides a representation of the causal processes involved in behavior change [329], though
other forms of domain knowledge can be a basis for interventions as well (e.g., disease etiology and epidemiology, theories from education and communication, etc). Advocates for incorporating such information into PHI development argue that it can inform both system design and evaluation [206], and several studies indicate that the application of behavioral theory is especially beneficial in informing the content and timing of provided feedback [107, 416].

Theory-driven designs are indeed believed to be more effective at changing people’s health behaviors and attitudes than atheoretical ones [329, 518]. A main reason is because interventions that build on existing theoretical knowledge are better able to account for the factors central to behavior change, which enables intervention efficacy to be optimized [327]. Systems informed by theory are also more effective since they are better able to address the multiple and often unintuitive personal barriers to behavior change [259]. In addition, theory-based interventions allow the developer to avoid design choices based on assumptions that may not only lack evidence but have even been invalidated [329]. Further, theory-driven approaches that target specific data might better manage and protect users’ privacy concerns.

Finally, using behavior change interventions based on theory is advantageous because that theory then provides an explanation of why and how the intervention works (or does not work). This understanding can in turn facilitate the development of better theories — and subsequently better interventions, again better theories, and so on [329].

Thus, when aiming to help a person change or maintain behavior, looking to domain knowledge helps a researcher evaluate what is currently happening with that individual (e.g., his or her health status, condition severity, behavioral
tendencies), why that may be happening (e.g., personal, environmental, or social factors), and what is likely to happen next — and how technology might intervene accordingly to bring about positive change. Altogether, incorporating domain knowledge can therefore enable the development of better systems, better theories, and ultimately, better health outcomes and broad impacts [206].

2.3.3 Examples of Domain-Driven PHI

As mentioned, most research in the PHI context has considered behavioral theory as the primary source of domain knowledge. Researchers draw on this information mainly to inform interface design, determine target users, and guide evaluation [206].

Prominent examples of theories commonly employed to inspire and inform behavior change technology include goal-setting theory, the transtheoretical model, and Fogg’s behavioral models for persuasive design. These are indeed well-suited to the task. Goal-setting, the intentions behind goals, and goal-directed activities have been theorized as integral to behavior change and the attainment of said goals [28, 29, 188, 294]. Empirical research further confirms goal-setting as a key behavior change technique [328]. The transtheoretical model (TTM) [403] then provides a conceptual framework to evaluate a person’s readiness to embark on a goal and monitor her progress through stages of behavior change (precontemplation, contemplation, preparation, action, maintenance, and potentially relapse).

One of the earliest and perhaps most well-known examples of a PHI technology guided by such theories is the UbiFit system for monitoring and main-
taining physical activity [107]. Influenced by the TTM, UbiFit targets the contemplation, preparation, and action stages of change [107, 108] and incorporates goal-setting theory [295] into design choices. Fish’n’Steps [290] also drew inspiration from the TTM and goal-setting theory in its similar implementation of a personal display that reflects a user’s physical activity levels through fish characters whose size and facial expressions map to goal progress and achievement.

Other PHI examples designed on premises of goal-setting include systems to encourage physical activity [349, 350] including at the group or family level [102] or in combination with other wellness activities such as diet management and relaxation [177, 316]. Content analyses of health management apps provide additional examples of recent tools implementing goal-setting techniques [302].

Some systems attempt to help users develop positive habits related to their goals [386, 466] by incorporating psychological theory related to habit formation [459, 530] or positive and negative reinforcement [326]. Other systems that similarly incorporate reward and punishment schemes based on reinforcement theory often use gamification elements to motivate user engagement [370], though such approaches have met criticisms for shallowly applying those theories or using methods that border on exploitation [53].

Lastly, Fogg has supplied a series of guidelines to promote behavior change and corresponding design recommendations for persuasive technology [162, 164]. Fogg’s Behavior Model for Persuasive Design (FBM) [165] is comprised of three elements: motivation (e.g., hope, fear, social acceptance or rejection), ability (e.g., time, money, effort), and triggers (e.g., inspiring videos, reminders), which together determine whether a target behavior will be achieved. Persuasive technologies guided by the Fogg behavioral theories span multiple do-
mains related to health (e.g., physical activity [7]) and beyond (e.g., sustainability [472] or social activism [411]).

Aside from these commonly leveraged frameworks, other theories from psychology and behavioral science are sometimes utilized by HCI researchers too. For instance, UbiFit also used Goffman’s theory of Presentation of Self in Everyday Life [185] to inform design choices about the types of control to give a user over personal information [107]. However, as mentioned, most PHI systems focus on goal-setting, TTM, or the Fogg theories, if any theory is adopted at all. This actually suggests an important open question about the extent to which such systems’ developers have genuinely determined that these theories are the most relevant for their work — or whether they are simply part of a “theory cascade”, selecting familiar, in vogue, and commonly employed theories rather than seeking out alternative ones that might in fact be better suited to their particular PHI projects.

2.3.4 A Domain–Practice Gap in HCI

Thus while some PHI technologies do draw on the science of behavior change, reviews find that the majority of health applications do not incorporate salient theory into designs [23, 206, 259, 368]. They also typically overlook other important sources of domain knowledge such as evidence-based clinical practice guidelines, to which reviews find few PHI tools adhere [5, 302]. This means that PHI work is often disconnected from valuable domain knowledge that could help ensure support is being provided in a clinically, contextually, and personally appropriate manner.
Such a domain-disconnected approach can constrain both how a system analyses a user’s personal data as well as the interventions it provides. First, consider analysis, the component of PHI technologies explained in section 2.2.2 that transforms personal data into health-related information (e.g., quantitative health metrics or qualitative behavioral feedback that can be reflected or acted upon). To compute such metrics or build models about a user’s health, PHI systems many times use statistical or machine learning methods. However, instead of defining variables or constructing features in a manner grounded in domain knowledge (e.g., by operationalizing well-established theoretical constructs), many systems take a more data-driven, “black box” approach (i.e., capture an abundance of data, run a model, and let the patterns shake out).

One problem with this strategy is that it can produce noisy models [34] that are inaccurate [546] and do not surface important underlying relationships present in the data [196]. Researchers expressing skepticism about the reliability of these domain-disconnected measurements argue that analytic outputs may be incomplete or even misleading [300, 301]. Further, scholars have pointed out other troubling human-centered aspects to data-driven user modeling, since it reduces the user to a metric to be algorithmically optimized, relies on a narrow set of assumptions regarding who the user is, and overall cannot fully accommodate the diverse, multifaceted nature of human identities [278, 320].

Therefore, these systems may not be modeling the factors most relevant to the health outcomes of interest nor adequately supporting the idiosyncratic user who will engage with the output of such PHI analyses. And though these systems may be capable of identifying statistical patterns in data logs, without a theoretical foundation, it is easy to misinterpret such observations or fail
to control for confounding factors underlying those results. Said another way, domain-disconnected systems may simply have a veneer of robustness, where success has been achieved in robustly fitting a model — but this display of validity holds little practical value if the appropriate health determinants are not being targeted in the first place. Instead, using domain knowledge can improve the relevance of variables analyzed, particularly to ensure that analysis does not omit factors known to be central to behavior change (e.g., relevant psychological processes, personal traits, contextual information, etc.) Ultimately, this can improve the utility of models and the interventions they drive [196, 327].

For example, Health Mashups, as mentioned, extracts patterns from various sensed and self-reported data streams to help a user answer questions like, “Do I sleep better on nights after I work out?” [45]. However, daylight exposure — something well studied in the chronobiology domain — can have a substantial impact on sleep, even more so than exercise [60]. In this example, the system would indeed be useful in helping an individual discover links between sleep and exercise; but if that person were aiming to improve her sleep, it would actually be preferable to target lifestyle changes related to daylight exposure, which a domain-driven strategy would be equipped to handle.

In addition, a common data-driven rationalization is that the pathway to better modeling is simply more data (e.g., in the above example, this would go something like: if we had captured more and more data, including information like daylight exposure, the model would have eventually been able to discover its strong impact on sleep, even if a domain-informed strategy had not been followed). However, this is not really a practical solution, as such exhaustive approaches are not particularly feasible for either humans or machines [292]. First,
it is unrealistic to expect an individual to manually capture such an expansive amount of data. This task is difficult for technology as well, due to physical and practical limits of computing. For instance, battery drain can plague even “simple” continuous sensing toolkits [274, 515]. Similarly, algorithms for data cleaning, aggregation, and compression are still not up to the task of processing or storing the vast amount of data personal health devices can generate (e.g., a clinical heart rate monitor produces approximately nine gigabytes of data in a month) [337, 476]. Domain knowledge can help avoid unnecessary processing that uses up finite computational resources, while ensuring important facets of data are not overlooked.

Beyond the analysis phase, failing to incorporate domain knowledge into the design of a PHI system’s interventions is similarly problematic, considering the benefits of domain-driven strategies overviewed earlier in Section 2.3.2. But while theoretically-informed interventions are more successful at producing positive outcomes [328, 417], reviews and content analyses (conducted by other researchers as well as myself) find that many system designers do not draw on theory and instead rely on intuition or trial and error [373]. For example, digital smoking cessation tools rarely follow established clinical practice guidelines for treating tobacco use and dependence, plus these tools are typically one-size-fits-all and use gameified elements to maintain engagement, rather than providing self-help strategies that are tailored to the individual characteristics (e.g., motivations, barriers, pace of goal progression) that behavioral theory has established are key predictors of cessation outcomes [355]. Similarly, PHI technologies aimed at supporting healthy sleep tend to encourage behavioral modifications based on folk wisdom, rather than providing personalized support that accounts for factors scientifically-backed as relevant to sleep [3].
Even in cases where domain knowledge does play a role, it is often used only to explain behavior but not to change behavior (i.e., to inform interventions) [329]; or, systems that claim to incorporate theory into delivered interventions often only give theory a cursory mention and leave details ambiguous as to how constructs were actually translated into design elements [259, 382]. As a result, some researchers question whether these personal health informatics systems have the ability to produce measurable, long-term behavior change [303, 317].

2.4 A Framework for Domain-Driven PHI Development

As just described, system designers taking a domain-disconnected approach may not build in support to target the most theoretically, clinically, or personally meaningful health determinants for assessment and intervention. A greater awareness of domain knowledge can enhance the creation of PHI systems across all phases of development, by informing choices regarding how to collect data, model health, and provide feedback in ways that will help people more effectively self-manage their wellness.

It is therefore vital that HCI’s PHI community adopt more domain-driven approaches that base development decisions on foundational knowledge from disciplines relevant to the aspects of health on which a given technology is focused. A main challenge in doing so, however, is that the field lacks an explicit set of guidelines for how to actually go about domain-driven development.
In this dissertation, I provide a framework that specifies just such details. In doing so, I pursue a long-term vision of end-to-end PHI systems that use domain-driven approaches to capture and analyze a variety of pertinent signals from personal data streams and that implement domain-aware design guidelines for providing tailored feedback about the self-manageable factors significant to health.

Figure 2.1 illustrates this framework — the reusable development pattern I have used to conduct the PHI research presented in this dissertation. Each subsequent chapter more fully describes and demonstrates each of these framework components, which at a high-level are as follows:

- **DOMAIN INQUIRY.** The first step in conducting domain-driven development involves identifying an application area where personal health technology could have meaningful impacts, considering the feasibility and appropriateness of such a technology-based solution, and determining the relevant domain(s) from which knowledge can be drawn. Rich bodies of scientific literature or empirical evidence may hold such knowledge. For example, I draw upon years of chronobiology research to support my case study research on sleep, daily performance, and emotional wellness. Domain inquiry can also involve engaging with domain experts or learning about the lived experiences, extant practices, and perceived needs of users. Altogether, gleaned domain knowledge can provide a deep understanding and sensitivity about the role of technology in a given health context, which can inform subsequent development steps.
• **DOMAIN-DRIVEN HEALTH ASSESSMENT.** Knowledge gathered during domain inquiry reveals valuable constructs to operationalize, suggests types and sources of data likely to hold information relevant to those constructs, helps guide analysis of that data once captured, and aids in interpreting analytic results — all activities undertaken in this phase of domain-driven health assessment. Compared to more purely data-driven approaches, using domain knowledge to drive variable selection reduces the chance of emphasizing convenient items that might not be relevant and helps reduce the computational costs that come with calculating extraneous features. Domain-driven approaches are also often more interpretable; knowing that a particular variable is informative has some value, but understanding why it matters can help modelers, designers, and theorists choose variables that are more likely to be appropriate to their needs. I emphasize this in my case study research, where I strive to present findings in a way that goes beyond describing *what* was observed to get closer to *why*. Further, such insights can fuel subsequent design work by suggesting effective interventions that target significant health determinants.

• **DOMAIN-AWARE INTERVENTION DESIGN.** An overarching goal of this domain-driven framework is to support health self-management with end-user tools. In this iterative phase of intervention design, guidelines are determined regarding what feedback to present and how, mockups and prototypes of various fidelities are built, and user models produced during the framework’s phase of assessment are instantiated in personalized interfaces. Throughout this cycle, participatory activities support the evaluation and continued refinement of these domain-aware, user-centered designs.
Figure 2.1: A framework for domain-driven PHI development, comprised of stages for tapping domain knowledge, collecting and analyzing personal data to assess health, and designing user-facing feedback.
This framework provides a footing for an HCI researcher undertaking a domain-driven PHI endeavor in any sector (e.g., academia, industry, government). As such, these guidelines are intended to be a flexible “process guide” rather than a binding set of cookbook practices. These strategies may not be applicable literatim for every PHI solution, but my research presented in this dissertation shows instances where they have worked successfully. The bulk of my work is from a case study on bringing domain knowledge from the field of chronobiology to the design of PHI technologies that support healthy sleep, cognitive performance, and emotional wellness. This chronobiology-informed perspective has allowed me to develop approaches to computationally assess sleep-wake behaviors, disruptions, and daily alertness levels. In addition, I have used this approach to guide design work for technologies that deliver personalized, biologically-aware feedback to improve sleep habits, productivity, and mental health.

To conclude this chapter, I will reiterate that HCI researchers are increasingly exploring ways technology can support personal health — work that has made swift advances and imparted legitimate benefits to people striving to gain personal insights, manage conditions, and improve overall wellbeing. However, this field is still young, and it continues to change with the regular arrival of new devices, applications, and interfaces born in both academia and industry. It also faces challenges related to the lack of explicit, well-established, and domain-informed procedures for analyzing personal data or delivering effective interventions. I argue that taking domain-driven approaches can help to realize the full potential of these tools for self-managing health, which would lead to more successful experiences with technology and more positive health outcomes overall.
The first step in developing a successful Personal Health Informatics (PHI) system is gaining knowledge in the domains relevant to that endeavor. As defined in Section 2.3.1, I characterize domains as areas of understanding that can be brought to bear on PHI research, design, and development. Scientific theories, methodologies, and empirical evidence could all be considered types of domain knowledge, as could practitioner expertise and an understanding of problems and extant solutions within that space. I refer to the practice of gaining this understanding as domain inquiry. (See “Domain Inquiry” portion of Figure 2.1).

Domain inquiry is not a practice uniquely useful for creating PHI technology. A software architect would not likely attempt to build, for instance, a distributed banking application or any other non-trivial system without first ensuring enough familiarity with relevant business areas. Having an ability to make good development choices — whether about banking products or effective health management tools — results from having sufficient understanding of salient domains [372]. This chapter identifies and demonstrates practices to attain that understanding.

I consider the following as core components of the domain inquiry process: (a) selecting an application area and relevant domains, (b) gathering knowledge from said domains, and (c) using this knowledge to plan subsequent stages of PHI work. In this chapter’s introductory section (3.1), I describe each inquiry component in turn, at a high level. While I provide a few brief examples for illustrative purposes, I keep these initial descriptions at a more general, conceptual level to make it easier for a reader to find and scan a boiled down set of
inquiry guidelines. Then, to demonstrate each inquiry component in practice (and as a way of giving the reader the specific, prerequisite background about my case study work), each following section of this chapter (Sections 3.2, 3.3, and 3.4) revisits the inquiry components and reports on the outputs of the inquiry process for this dissertation’s case study on sleep, cognitive performance, and emotional wellness.

3.1 An Overview of the Inquiry Process

Selecting an Application Area

The initial phase of domain inquiry involves identifying a compelling problem area where a PHI solution would be valuable, followed by determining the area or areas of expertise (i.e., domains) that are useful for informing the work to develop that technology.

In selecting an application area, several factors can be important to consider. Again, these principles are not necessarily exclusive to PHI — most funding agencies’ review guidelines would likely be along similar lines (e.g., the National Science Foundation’s evaluation criteria for “Intellectual Merit” and “Broader Impacts”). Rather, these are well-established motivations I feel are important to articulate here to encourage their contemplation.

Perhaps foremost, it is valuable to think about the potential for the research to deliver significant benefits on individual, group, and larger community levels. In a nutshell: is the research worth doing? These benefits may easily align with well-established societal goals (e.g., improving human health or quality of life,
preserving the environment, or empowering users especially from underserved or vulnerable populations). Though, an issue does not have to be grandiose to merit attention — modest research goals can certainly be worthwhile too.

Another element to consider is the potential of work in that area to meaningfully advance scientific knowledge and understanding about that topic. Such understanding can in turn be practically applied or used to devise solutions to problems (e.g., the current PHI problem being addressed or problems that are the focus of future work in the area). A researcher may also want to keep in mind whether an application area is well-suited to interdisciplinary study (i.e., that can transcend the boundaries of HCI and the target domains). For example, the application of computing to health and medicine is a burgeoning, integrative research area — one where future research activities could continue to cross additional disciplinary boundaries, for instance, by connecting with geographic information science to study various issues central to the spread of diseases.

When selecting an application area, it also makes practical and potentially ethical sense to determine whether a technological solution is actually feasible as well as appropriate in that context. As an example of infeasibility, a PHI approach dependent on a well-developed internet infrastructure would not be deployable nor cost-effective in some rural areas [538]. As far as determining the appropriateness of a technological approach, an example from my own work on developing technology for managing bipolar disorder illustrates the importance of asking this question — and the difficulty that can come with answering it. Namely, I found that patients can experience a number of positive outcomes from using technology for condition management but that the same technology can sometimes have an agitating effect on symptoms or even trigger a relapse —
a tension requiring careful navigation as a system designer [354] and discussed further in Section 6.2.1. Such tradeoffs should be pondered at this point; and if it is determined that the injection of technology in a context would come with practical obstacles, ethical risks, or other drawbacks likely to outweigh potential benefits, then perfectly legitimate outcomes of this stage can include seeking out a different application area or exiting the inquiry process entirely (and perhaps disseminating these discovered implications to not design technology here [38]).

However, if a PHI solution is deemed advantageous, we move on to identifying salient sources of domain knowledge.

**Identifying Salient Domains**

After selecting an application area, we can next determine relevant domains from which to draw background knowledge. Such information is highly valuable, as it provides a lens through which to study the topic of interest, increases a researcher’s awareness of existing challenges and solutions in the space, and helps scope and direct the work. This deeper understanding also suggests befitting methods, guides potential analyses, and aids interpretation of findings. Overall, domain knowledge thus helps a researcher more thoughtfully reflect upon a problem and offers practical advantages as well.

One way to initially go about identifying pertinent domains is by contemplating the goals of the research. For example, the goal of the aforementioned UbiFit system, a healthy lifestyle intervention technology, was to motivate individuals to change their behavior (specifically, do regular physical activity) and sustain those changes [105]. UbiFit’s designers therefore looked to theories from behavioral psychology that focus on behavior change [107]. Though a number
of such theories exist, the researchers identified two theories as particularly relevant to their research goal: Locke and Latham’s goal-setting theory [295] and Prochaska’s transtheoretical model of behavior change [402].

My own research on smoking cessation [355] provides another example. In this case, I leveraged social media data to predict a smoker’s likelihood of successfully quitting, and my ultimate goal was to develop a PHI system that could use such predictive models to provide tailored behavior change support. So, like UbiFit’s designers, I sought out behavior change theories. In my review of the literature, the transtheoretical model (TTM) stood out, especially because numerous studies over the years have tested it with smoking cessation interventions, with findings generally supporting the model and confirming that it can be used to predict smoking abstinence [129, 404], including for a diverse set of smokers with various demographic backgrounds or smoking behaviors [419, 500]. The choice worked well; the TTM helped guide my assessment phase, providing me a conceptual framework with which to first evaluate an individual’s readiness to embark on a cessation goal and then monitor her progress through stages of behavior change. Given their apparent utility, these social media based measures that captured meaningful aspects of TTM could then be embedded into an envisioned tool for smoking-related behavior change.

A final example is provided by this dissertation’s primary case study, which focuses on sleep, cognitive performance, and emotional wellness. These aspects of health are essentially biological processes, so I looked to biology for domain knowledge. More specifically, my reviews of the literature and consultations with medical experts led to the branch of biology known as chronobiology, the field of study concerned with the rhythms that guide biological functioning.
Gathering Domain Knowledge

Having identified a compelling application area and relevant domains that can inform the PHI work, we move on to gaining a deep understanding of those domains. Essentially, this involves eliciting knowledge from various sources — for instance, through continued review of the literature and conferral with domain experts (e.g., clinicians, chronobiologists, etc.) as well as by engaging with envisioned users to ensure a system meets their needs and respects their extant practices.

Such processes can be time-consuming and challenging. Identifying, contacting, and meeting with experts who may be remotely located and available in short supply can be particularly difficult. Still, such efforts are highly worthwhile to increase the chances that future steps move in the right directions and that PHI outcomes will be efficacious (e.g., clinically relevant and personally beneficial). And for many contexts, these efforts are not just advantageous but imperative. Given PHI’s focus on health and especially for applications for vulnerable populations, ill-informed assessment and design work does not only risk irrelevance or poor usability but can have lethal consequences, not to be too dramatic. For example, delivering a sleep plan unsuited to an individual’s biological profile can contribute to circadian disruption, which (as I will describe later in this chapter) is associated with numerous negative health consequences; or a misguided tool for managing a mental health condition like bipolar disorder could trigger a life-threatening relapse.

Overall, gathered domain knowledge can provide an informed worldview from which to make thoughtful decisions during a PHI enterprise [372], including to inform the work’s scope, modeling strategies, and design requirements.
Informing PHI Development

Having acquired background information and a strong foothold of understanding, we can devise an informed strategy for assessing and supporting health. First, domain inquiry helps define the scope of this PHI work. Especially if domain knowledge suggests that an area is associated with a high degree of intra- or inter-individual variability, then it will be challenging to develop a PHI solution that meets such manifold needs within and across people and situations. Instead, it may make more practical and impactful sense to initially concentrate on a scoped subset of the population that is particularly at risk or a select aspect of a condition that is a strong determinant of overall health. Relatedly, it is important to consider here whether a personalized PHI solution is necessary or whether something more generic would suffice or even be preferable for a given context. Knowledge about individual differences in a pertinent domain can therefore further help in determining whether data can be analyzed at an aggregate level or whether it is necessary to obtain per-person information and develop individualized models.

Next, operationalizing constructs, defining variables, and selecting a methodology can all benefit from a researcher’s awareness of domain knowledge, which helps ensure salient concepts in an application area get appropriately translated into measurable factors. Again using my work on smoking cessation as an example, domain knowledge from behavioral psychology helped me realize the strong influence that an individual’s motives, mindsets, and strategies could have on the outcome of a cessation goal [295, 402, 403]. Based on such information, I defined “Behavior Change Process” variables designed to model a person’s cessation process and evaluate whether or not she was exhibiting behavioral signals known to correlate with successful outcomes.
Analyzing these domain-driven variables produced striking results. I found that individuals who relapsed were far more likely to: quit for more casual, shallow, and unrealistic reasons; procrastinate before cessation (procrastination is a known marker of low commitment and unsuccessful outcomes in the behavior change literature [129, 402]); and choose a cold turkey strategy rather than use more effective treatment methods that I discovered during domain inquiry.

Finally, domain knowledge helps in identifying data likely to hold information relevant to these constructs and in guiding the capture of this data. As described in Section 2.2.1, personal data can be collected in several ways — manually or passively, via hard or soft sensing, or through some combination thereof — about a variety of physiological, psychological, behavioral, or contextual information. Domain knowledge can help in making choices about which strategy to take and which data to acquire, depending on the goals of the PHI project. To use my smoking cessation work as an example one last time, domain knowledge helped me identify that the tweets and social network information of smokers on Twitter were an attractive source of data for assessing key personal variables from the domain literature (e.g., preparedness, emotional distress, temptation exposure, social support, etc.) that a cessation intervention tool would want to harness in predicting relapse and providing tailored behavior change support. My strategy was also inspired by that of a growing body of recent research focused on conducting “natural experiments” [131] on social media, which have demonstrated that the abundance of data on such sites creates a scenario that resembles the environment of a traditional controlled experiment and can even support causal discovery [380] during the stage of assessment.
In the following sections, I return to each component of domain inquiry (selecting an application area and identifying salient domains, gathering knowledge from those domains, and using that knowledge to plan strategies for assessment and intervention) to demonstrate them in action using my case study on sleep, performance, and emotional wellness. Furthermore, by reporting on and documenting the outputs of my domain inquiry process in this context, I hope that these sections will also supply constructive motivation and background information for chronobiology-driven PHI research and serve as a useful knowledge resource for readers interested in doing future work in this area.

3.2 Selecting an Application Area

To begin, I explain this case study’s motivations to focus on sleep and its impacts on physical, cognitive, and psychological functioning. Specifically, in the subsections that follow, I explain why I believe this is a compelling problem area that is ripe for novel technological solutions, I review related work already underway, and I point out how domain knowledge can help to address shortcomings with these extant approaches.

3.2.1 Sleep, Cognitive Performance, and Emotional Wellness

Sleep plays a pivotal role in our overall health. It has a direct and substantial impact on numerous aspects of our daily lives, from the functioning of our immune system to our decision making abilities to our psychological wellbeing — and even moderate sleep disturbance can have severe detrimental effects [277].
A large body of scientific evidence links poor sleep habits (e.g., too little, too much, or erratic sleep [340]) to a range of negative consequences for physical health, including obesity and poor dietary habits [16], greater susceptibility to illnesses like the common cold [100], and an increased risk of more serious, chronic diseases such as diabetes [192, 540] and heart disease [22, 83]. Generally speaking, persistent sleep loss including from clinical sleep disorders (e.g., insomnia) is often comorbid with and indicative of numerous other debilitating health conditions [155, 166, 306, 387].

Sleep also has well-known effects on cognitive performance [184]. Sufficient, high-quality sleep increases productivity and work performance [103, 432]; enhances learning and problem solving [509]; and improves energy, alertness, and reaction time [232, 428]. On the other hand, inadequate sleep can suppress brain function, interfere with memory consolidation [308, 509], impede learning, and hinder concentration [56, 132].

In addition to directly impairing performance, inadequate sleep is also associated with subsequent feelings of fatigue [268, 269]. Such reduction in alertness and functional ability can pose a serious safety hazard, as it significantly increases the risk of industrial and motor vehicle accidents [57, 132, 477]. In fact, it is estimated that over a third of all road accident fatalities in the United States result from driver fatigue [280], and statistics are similar for fatigue-related aviation and marine accidents [367].

Beyond the physical and cognitive effects, insufficient or erratic sleep is a major risk factor for developing various psychological problems, including depression [167, 293], anxiety [180, 191], and stress [30, 334]. Sleep disturbance may also trigger symptoms or even the onset of mood disorders like bipolar
disorder [189, 399] and schizophrenia [395]. In addition, poor sleep is one of the top factors contributing to overall unhappiness [127].

Such mental health issues affect a significant portion of the world’s population and can result in debilitating and life-threatening outcomes. As such, mental health is becoming an increasingly pressing health care issue. Globally, about 450 million people suffer from mental illnesses, and neuropsychiatric disorders make up 4 of the 6 leading causes of years lived with disability [532] — though, prevalence and impact may be even greater, as many cases go undiagnosed and the burden of psychiatric conditions is likely heavily underestimated [298]. These problems are particularly acute for younger generations, for whom inadequate sleep is widespread and associated with stress, suicidal thoughts and actions, and depression [65, 208, 254, 481, 514, 529] — emotional problems reported as so severe by a majority of students that they regularly impact daily functioning [231]. Finally, affective health ties back to physical health, with negative affect linked to an increased risk for illness and mortality and positive affect linked to lower morbidity and better health outcomes overall [101, 468].

An ample amount of quality sleep is thus essential for maintaining physical, cognitive, and emotional health. However, despite evidence establishing the importance of sleep as well as public health organizations’ efforts to promote healthy sleep routines, chronic sleep deprivation is prevalent [227]. Sleep pathologies and associated conditions resulting from poor sleep are considered to be reaching epidemic levels, affecting millions of people around the world [57, 306]. According to the U.S. Centers for Disease Control (CDC) and the National Institutes of Health (NIH), sleep disorders affect 50–70 million people in the U.S. alone, with many others likely not yet diagnosed [103].
According to the U.S. National Sleep Foundation’s latest report, insufficient or poor sleep negatively impacts daily activities at least once a week for 45% of Americans [366]. Duration, consistency, and quality of sleep could all be improved. While it is recommended that adults get at least 7–9 hours of sleep per night and younger individuals another hour or more beyond that [210], people typically report getting an average of between 6.5–7.5 hours [365, 366]. In addition, maintaining a consistent sleep schedule can be as or more important for health than sleep duration [448], yet the same studies find irregularity in sleep-wake behaviors, with duration fluctuating an average of 40 minutes between work and free days. Further, over a third of questioned individuals report that their sleep quality is "poor" or "only fair" [366].

Practically speaking, the annual direct and indirect expenses of treating sleep-related problems are estimated at $14 billion and $150 billion, respectively [103, 512]. Mental health problems are associated with substantial financial costs too — over $100 billion annually for mood disorders like anxiety and depression [248]. The intangible costs of physical, cognitive, and emotional difficulties can be far greater than any dollar amount, however, on both personal and societal levels [347].

Poor sleep and its consequences have thus been identified as a crucial challenge that calls for researchers, including those from the HCI community, to seek solutions in a multi-disciplinary effort [277].
3.2.2 Opportunities for Technology

I argue that technology can help address these challenges, for reasons similar to those I presented throughout the introductory sections of this dissertation. Technologies (e.g., personal computing devices, the internet, and the digital trails left from interacting with any of those systems) offer a tremendous opportunity to study, monitor, and positively modify behaviors. In particular, technology seems well-suited to helping people achieve better sleep, optimize cognitive performance, and manage emotional wellness, for the following reasons.

First, the fact that personal technologies are carried and used in naturalistic settings, throughout the course of one’s daily activities, and by a large and diverse population of people means that they provide a window through which we can study phenomena of interest in a broad, unobtrusive, and affordable way. This ability to continuously analyze connections among experiences, behavioral patterns, and health indicators can help us gain a greater scientific understanding about these subjects — for example, by enabling long-term observation of how sleep quality relates to cognitive performance or by contributing to our fundamental conceptualization of mood disorders [524].

Second, technology-based monitoring can also support assessment. Symptoms of poor sleep often go unnoticed, both because most people do not undergo lab-based sleep tests that might help identify clinical disorders and because most people’s awareness is limited when it comes to their own sleep patterns and quality [502] or what constitutes healthy sleep practices in the first place [366]. Data from personal tracking tools could help doctors detect sleep or mood disorders that often go undiagnosed [444] or help individuals identify lifestyle factors negatively impacting their sleep and subsequent daily func-
tioning [92]. Passive approaches might work especially well to promote such outcomes while alleviating some of the burdens and biases of self-monitoring.

Finally, personal technology provides an accessible medium for delivering feedback and interventions that can guide an individual to improve sleep habits and manage cognitive performance and mental health. And importantly, potential users express a strong need and receptivity toward such tools [92].

It therefore seems compelling and appropriate to explore the use of technology to measure and manage sleep along with related aspects of cognitive and emotional health. One of the most comprehensive examinations of this topic to date is similarly optimistic, outlining a number of design opportunities in this space [92]. Numerous HCI researchers appear to agree (and perhaps thousands of consumer app developers, based on my recent searches of the Android and iPhone app stores [356]). Indeed, these areas of health have gained considerable recent interest from the HCI community in both academia and industry.

3.2.3 Related Academic and Commercial Work

Reviews of sleep technologies identify five main platforms: mobile devices (e.g., smartphones and tablets), wearables sensors worn directly on the body (e.g., bracelets and smart clothing), environmental sensors (e.g., mattress sensors and bedroom wall cameras), desktop or website platforms, and accessory appliances (e.g., specialized alarm clocks, wake-up lights, and sound machines) [257]. Such tools generally focus on measurement (i.e., tracking via manual input, automatic sensing, or a combination) and/or intervention (i.e., providing feedback about current sleep habits or recommendations for making improvements).
Some prominent and recent examples of research prototypes developed to help users manually record sleep data with mobile phones include The Wellness Diary [318], Sleepful [277], SleepTight [93], and ENTRAIN [510]. Commercial tools for manual sleep journaling exist as well, but fewer are available today than just a few years ago, in favor of more passive tracking apps. For example, apps like Tylenol SleepTracker or the once popular YawnLog are now defunct; while in contrast, the passive sensing apps ElectricSleep, Sleep As Android, and SleepCycle, which were among the first to attempt automatic sleep tracking using smartphone accelerometers, still have thousands of users on the Android and iPhone app stores.

Academic research has similarly been shifting toward the development of more automated and unobtrusive approaches. Much work centers attention on automatic sleep measurement using various smartphone sensors. For instance, the systems iSleep [200] and wakeNsmile [267] use a phone’s built-in microphone to detect sounds (e.g., snoring, coughing) and body movement in order to predict sleep phases, while ApneaApp [362] emits frequency-modulated sound signals from a phone to detect sleep events through a sonar-like system. Sleep-Miner [24], Best Effort Sleep (BES) [89], and Toss’n’Turn [333] have similarly used ambient sound and light together with phone usage data such as screen unlock events, battery status, app use, and communication logs to predict sleep stage, quality, and duration. Such phone use data along with contextual information like time and location have also been used to predict bedtime, sleep duration, and irregular sleep patterns [217] as well as sleep quality [225].

While not as prevalent as smartphones, wearable devices are seeing an increasing rate of adoption as affordability and accuracy improve, and most wear-
ables now on the market provide automatic forms of sleep measurement. Wearable devices designed to manage diagnosed sleep disorders such as sleep apnea have been around for years, but recent efforts have been made to design less invasive tools; for instance, WatchPAT [521] asks users to attach a probe to a finger during sleep so that the system can detect respiratory disturbance by monitoring peripheral arterial tone (PAT). Regarding more general sleep monitoring, early commercial wristbands such as the WakeMate and those by Lark Technologies used actimetry-based sensing, which is often used in clinical settings [12], to measure nightly sleep duration and quality. The Zeo headband, another early product for at-home sleep monitoring, provided an alternative wearable form factor and was particularly well-received by Quantified Self enthusiasts; however, the device was considered too cumbersome by the average consumer, and the company is no longer in business. To explore other body-based sensors, academic researchers have developed a prototype neck-cuff system for real-time sleep monitoring [430]. But today, the best selling wearables for sleep tracking are wrist-worn devices like the Fitbit, Jawbone, Apple Watch, and Microsoft Band, which use accelerometers to measure movement and determine sleep onset and wake times as well as phases of light and heavy sleep.

A main reason sleep assessment has been moving away from self-tracking and toward more automated methods is that manual tracking tools are often burdensome since they depend on users explicitly indicating when they go to bed and when they wake up day after day, and any inadherence hinders their utility. Using smartphone or wearable sensors for passive tracking helps relieve some burdens, but these approaches are not fully immune to such problems either, as they often require users to keep the phone in bed during sleep, can be intrusive or uncomfortable to wear, and face challenges to sensing accuracy.
introduced by sleeping partners and pets or a user’s failure to properly charge or configure the device.

Recent academic attention has therefore been placed on creating contactless sleep assessment systems that use environmental sensors. For instance, using an off-the-shelf Doppler radar sensor, DoppleSleep [409] monitors breathing rate, heart rate, and body motion in order to classify sleep states at an accuracy comparable to more intrusive and expensive clinical sleep sensing techniques. The Lullaby system [242] uses bedroom sensors and cameras to record temperature, sound, light, motion, and pictures in order to help users identify environmental factors responsible for interrupted sleep. Commercial contactless tools exist as well. For example, a bedside device by ResMed uses sonar to monitor a person’s breathing rate and associated sleep stage. Devices like Withings Aura, Beddit, and Sleepace Reston are similar to other tracking technology in that they use body motion as a way to assess sleep, but they use mattress-based rather than body-worn sensors. Along the same lines, the sleep tracker Sense clips to the user’s pillow.

Another approach has eliminated physical sensors entirely and instead focuses on leveraging technology usage patterns as a way to assess aspects of sleep or daily behavior (i.e., soft sensing, as described in Section 2.2.1). By applying computational techniques to naturalistic data, such work aims to infer health-related behaviors and mechanisms and in a manner that can be more broadly, cheaply, and quickly deployed than with a physical device like those described above [120]. Regarding sleep assessment, smartphone usage logs about app launching, outgoing and incoming communication, and screen unlocking have been used to predict sleep stages [333], duration [217], and qual-
ity [24] as mentioned earlier, though these data are often analyzed in conjunction with data from a phone’s hardware-based sensors (e.g., the GPS, battery, and ambient light sensors). Beyond smartphone-based usage, social media data (specifically, Twitter posts) have also been passively mined and used to detect individuals suffering from insomnia [224].

Also relevant to this dissertation, given that parts of its case study focus on cognitive performance and emotional wellness, are studies that connect aspects of cognition and mood with technology-mediated behaviors. Regarding the former, higher levels of attention have been linked to the use of particular types of computer and mobile apps (e.g., email, messaging, notification trays) as well as certain usage behaviors (e.g., window switching and lapses in device use) [309, 396], while inattention has been associated with short bursts of smartphone use [383]. Such usage behaviors along with demographic information and contextual data (e.g., time, location, light levels) have also been used to model boredom [309, 397] and proneness to boredom based on the types and amount of smartphone app use [311].

Similarly, other studies suggest mood and mental health can also be assessed using soft sensing, often using social media data in particular as a window through which to evaluate individuals’ psychological characteristics and affective symptoms. For example, behavioral cues detected in Twitter data have been used to identify individuals experiencing depression [388] and predict the onset of depression [123], and similar analyses have been applied to Facebook data in order to detect signs of postpartum depression in new mothers [122]. Beyond social media, aspects of internet use more broadly have also been used as a marker of mental health problems. For example, excessive video viewing
and late-night use have been associated with symptoms of depression for college students, while frequent email checking has been identified as a signal of high anxiety levels [263].

Personal sleep information — whether captured manually or passively and via physical hardware or soft sensors — can then be used to help users gain insights into their sleep-wake habits, guide them in taking corrective measures to improve these behaviors, and ultimately help individuals achieve improved wellness. Based on this premise, a number of systems have been designed to provide such feedback.

Specifically, most of the aforementioned sleep tracking technologies also provide user-facing views into that data. For example, after syncing collected data manually or automatically with a mobile device, products like the Fitbit, Jawbone, and Microsoft Band allow users to browse sleep information related to duration, onset and wake points, and entry and exit from sleep phases. These interfaces also encourage users to reflect on how certain daily behaviors such as caffeine consumption or water intake during the day may have impacted that night’s sleep. Some apps provide additional information; for instance, the Fitbit calculates a measure of sleep “efficiency” based on sleep duration and the amount of time a person takes to fall asleep after getting in bed, while the Jawbone UP allows users to set sleep goals and provides progress reports. These tools also sometimes incorporate social features, though sharable data is usually limited to fitness or other activity-related information rather than sleep. A few exceptions are BuddyClock [251], which infers sleeping state (awake, snoozing, asleep) based on the status of a user’s alarm clock and allows a user to share that information with a social network, as well as “Got Sleep?”, described below.
Finally, when sleep technology began gaining traction in the HCI community a few years ago, Choe et al. [92] offered a forward-looking summary of various other feedback-related design opportunities for technology to support healthy sleep behaviors. The ShutEye smartphone app [36] realizes some of those ideas through a glanceable wallpaper display that conveys how various activities (e.g., drinking caffeine, exercising) will affect that night’s sleep if done at the current time. Other ideas have been implemented in SleepTight [93], which provides two forms of feedback: sleep summaries (visualizations of sleep duration and quality over various time spans) and comparative relationships among tracked data (e.g., the average sleep duration or the amount of caffeine consumed on nights of good, neutral, and poor sleep) to help users identify how various factors contribute to sleep quality. Similarly, the prototype app “Got Sleep?” [438] provides natural language summaries about a user’s sleep, suggests sleep targets, and provides features for sharing sleep scores on social networks in an effort to motivate users through competition and playfulness.

3.2.4 Connecting with Chronobiology

These efforts toward sleep assessment and intervention are encouraging steps toward supporting users in monitoring and improving their sleep-related behaviors as well as increasing our scientific knowledge surrounding people’s real life sleep and daily activity. However, this work often makes incomplete assumptions, interpretations, and design decisions, I argue, due mostly to a lack of grounding in a highly relevant domain: chronobiology.
Within our bodies there are hundreds of biological clocks coordinated by a “master clock” in our brain. These body clocks control oscillations in our biological processes, which vary significantly, predictably, and idiosyncratically throughout the day [79]. These fluctuations affect when we sleep and influence nearly all other aspects of neurobehavioral functioning as well, from digestion to concentration to mood [266].

However, the related work I just reviewed generally does not consider such information when assessing sleep (e.g., it does not factor in the effect of light exposure, a key factor in “setting” our biological clock). Similarly, PHI technologies aimed at improving cognitive performance are typically designed on assumptions that our capabilities over the course of a day are steady or could be made steady, when in reality, human biochemistry dictates that our performance levels naturally rise and fall throughout the day [79]. To build effective solutions in this space, researchers must account for the fluctuating nature of performance and the behavioral, environmental, social, and biological factors driving those fluctuations — something that current work rarely does, according to chronobiologists [445].

While studies often report a time-of-day effect, the fact that they do not take any biological factors into consideration hinders both the range of analyses explored as well as the ability to offer deeper explanations as to why particular trends are observed. For example, the aforementioned studies connecting aspects of cognitive performance with technology-mediated behaviors (e.g., [309, 397]) have identified consistent patterns of technology use and have associated cognitive or psychological states such as boredom and inattention with particular usage behaviors. However, such work does not provide satisfying
biological explanations of these patterns. Studies on technology use and sleep similarly lack the domain grounding necessary to more holistically interpret observations in a way that bears in mind latent biological aspects. Research guided by a theoretical understanding of the biology behind sleep and daily behaviors would be more aware of the need to investigate biologically-rooted factors in order to glean novel insights into idiosyncratic use behaviors.

Moreover, biological idiosyncrasy necessitates personalized approaches to modeling and feedback. Biological rhythms display individual differences, with body clock types falling on a spectrum from early to late types. This means that the mindset behind the maxim “Early to bed and early to rise makes a man healthy, wealthy, and wise” is erroneous for the majority of the world’s population [426] — yet it is a commonly held one, and it is often embedded into design choices that encourage all users to adhere to schedules that are really only appropriate for early types (e.g., the Jawbone UP’s “Early to Bed” goals encourage users to adhere to earlier and earlier sleep times). As another example, there are biologically-based individual differences in the effects caffeine has on sleep [542], but current systems tend to present more generic feedback (e.g., “End caffeine consumption 8–14 hours before bedtime” [36]).

Overall, more chronobiology-driven PHI work is thus desirable for several reasons. Studies with a restricted theoretical understanding of sleep and waking behavior provide a fragmented picture of both sleep and our broader daily experiences; instead, a grounding in domain knowledge would enable more holistic modeling of sleep and daily behaviors. Further, even though technology has the potential to improve sleep, technology can also impair sleep if not implemented properly [257]. In particular, research without biological under-
pinnings is unaware of the extent of individual differences in this context and, in turn, the level of personalization necessary. In addition, interventions that commonly target only sleep disturbances may merely be treating the symptoms of misaligned biological clocks rather than helping to address the root causes or work in tune with a user’s unique biological profile in the first place. A greater awareness of our innate biological rhythms could therefore positively impact how we design technology, enabling us to build more effective and personalized user-facing tools for supporting sleep, performance, and overall wellbeing on a broadly deployable scale.

3.3 Gathering Domain Knowledge

As explained at the beginning of this chapter, once a fertile application area and relevant domains have been determined, knowledge from those domains can be gathered from literature, experts, or any other identified sources that may contribute information that can strengthen the theoretical and methodological foundations of the research at hand.

In the following subsections, I present background on chronobiology, the domain providing that foundation for this dissertation’s main case study. Beyond chronicling my application of the domain-driven framework, I hope this section on gathered knowledge can serve as a standalone resource that is useful for others taking on chronobiology-driven work.
3.3.1 Chronobiology & Circadian Rhythms

As mentioned, chronobiology is the field of study concerned with the rhythms that guide biological functioning. Like that of nearly every terrestrial organism, human physiology has adapted to the periodic changes in sunlight exposure and temperature that occur as the Earth rotates around its axis approximately every 24 hours. Over this course of a day, our biochemistry varies significantly, causing regular changes in, for example, blood pressure, cortisol, and melatonin levels [428]. These changes follow what are known as *circadian rhythms*, a term that refers broadly to any self-sustaining diurnal biological cycle that keeps a roughly 24 hour period (“circa”: about, “diem”: a day) [197].

These fluctuations affect when we sleep, eat, and have an impact on our physical and mental performance and mood [266] — such as when we can swim the fastest (in the late evening) [39], when we are most prone to heart attack (in the morning) [348], when working memory has more capacity (generally in the afternoon) [79], and when depressive symptoms worsen (early morning) [526].

Jürgen Aschoff, a co-founder of chronobiology and the first researcher to investigate circadian rhythms in human beings, noted that “whatever physiological variables we measure, we usually find that there is a maximum value at one time of day and minimum value at another” [19]. His research introduced a new basis for explaining these patterns, namely by identifying that genetic determinants of behavioral rhythmicity (“clock genes”) are modulated by environmental information in order to keep our body clocks running [20].

More specifically, clock genes interact with each other to generate oscillations in gene expression. This successive gene activation forms a cycle, with the
initial activation of a gene regulated by the last gene in the sequence, creating an auto-regulatory feedback loop that takes about 24 hours [8]. Preciseness of this rhythm is maintained by a process known as “entrainment”, whereby a group of nerve cells in the brain use external information (predominantly sunlight) to keep our body clocks synchronized with changes in our environment [421].

The biochemical processes responsible for sleep and wake activity are influenced by two mechanisms working against each other: the body’s circadian oscillator promotes wakefulness throughout the day and determines the timing of sleep, while its homeostatic system increases the need for sleep the longer a person is awake and determines sleep duration, with sleep need abating during sleep [59, 127]. Social factors such as relationships and work schedules further affect our sleep patterns [422]. The timing and quality of sleep are thus influenced by three complex factors: our circadian rhythms, our homeostatic sleep drive, and a “social clock” based on social constraints [428].

Beyond sleep, biological clocks also influence our cognitive performance levels, which naturally rise and fall throughout the day [79]. Alertness, attention, reaction time, response inhibition, short-term and working memory, and higher executive skills all follow rhythmic patterns [49]. In my case study research, I focus on alertness in particular because it is considered a cornerstone of cognitive performance [445], correlates with a number of cognitive functions [14], and displays substantial variation over the course of a day [79]. In addition, alertness deteriorates considerably after lost and interrupted sleep [377]. As fatigue and the need for sleep accumulate while awake, an accompanying decrease occurs in cognitive ability and alertness [49]. These effects can become severe. For shift workers, the increased chance of accidents and injury due to
fatigue is well established [418]. More generally, the impairment effects of fatigue coupled with the endogenous decrements in cognitive functioning over the day have been equated to alcohol intoxication [269]. Fatigue also hinders meta-cognition and one’s ability to self-assess and recognize performance reductions [139], which may lead people to rationalize the sacrifice of sleep and disregard the well-studied negative impacts of sleep loss on performance, further compounding fatigue-based performance losses [183].

3.3.2 Chronotype

Our body clocks thus produce temporal fluctuations across a range of biological variables. In addition, these biological rhythms vary between individuals. That is, humans display inter-individual differences in the phase and amplitude of their circadian rhythms, from the timing of sleep-promoting hormone secretions [426], to the duration of sleep necessary to support health both short and long term [157, 266, 429], to the times when alertness peaks and dips each day [212].

A person’s chronotype reflects his or her unique circadian profile [109] that manifests in such biological and behavioral differences. A common distinction is made between early and late chronotypes (“early birds” and “night owls”) — people whose biological clocks drive them to sleep and wake earlier or later, respectively; however, chronotype is not binary but rather lies on a continuous spectrum from extreme early to extreme late [426]. Chronotype is a phenotype, meaning that it results from a person’s genetics interacting with features of her environment like light exposure [428]. Research has found that 50% of chronotype features are heritable [505].
Demographic factors such as age, ethnicity, and gender might also influence chronotype [426]. Children (preschool-aged as well as school-aged) are generally early chronotypes [457], transition to increasingly later types during adolescence, and reach a maximum lateness around 20 years old. One’s chronotype then begins shifting earlier once again; and in general, people over 60 years old have an early chronotype. The shift to a later chronotype begins sooner for females than males, which is in accordance with the general biological phenomenon that females tend to mature earlier. This means that men are relatively later chronotypes compared to females of same age for most of adulthood [426], until their chronotypes coincide around age 50, the average age of menopause.

Chronotype can also be impacted by light exposure. Longer exposure to daylight can shift individuals toward a later chronotype [428]; specifically, spending more than two hours outside has been correlated with chronotype shifting an hour later [421]. Chronotype can therefore also vary according to a person’s geography, based on the variability of seasonal sunlight duration at a given latitude and longitude [330].

### 3.3.3 Circadian Disruption

As described, our bodies’ circadian system plays a crucial role in synchronizing our internal processes with each other and with our external environments. However, a number of factors, which I detail below, can disrupt an individual’s circadian rhythms and, in turn, various aspects of functioning from sleep-wake cycles to metabolism to cognitive ability to mood stability [429]. Unfortunately, such problems stemming from circadian disruption affect daily life for millions
of people [347]. Chronic circadian disruption is linked to increased mortality in animal studies; and while it is often confounded with sleep deprivation for humans, research associates detrimental cognitive, behavioral, and physiological consequences with even brief disruption [239]. Living out of sync with our individual circadian rhythms can thus not only make us feel fatigued in the morning or frustrated at work but can be legitimately harmful, with serious long-term consequences for our health and wellbeing [266].

Foremost, while we are biological creatures, we are simultaneously social beings. Our combination of biology and society is arguably what separates us most from other animals, with our social structures dictating, restricting, and altering our biological responses [478]. Unfortunately, these social constraints often work against our innate rhythms. Every day, our internal circadian timings experience interference from externally determined social factors such as work schedules and leisure engagements. As examples, standard work schedules may require a late type (whose body clock wants her to fall asleep and wake later) to use an alarm in order to rise during what is still the middle of the night for her, biologically speaking; while an early type may stay up later than she would naturally prefer due to evening social schedules shaped by late types [426] but then be driven by her biological clock to wake up early, even if the amount of sleep obtained was inadequate [425].

The result for many people is markedly different sleep and activity patterns on work days versus free days [428]. Given that these demands manifest in sleep and wake fluctuations comparable to jet lag, this discrepancy is referred to as “social jet lag” since the causes are socially rooted [528]. Unlike the transient misalignments of jet lag from travel, however, social jet lag can be chronic.
Earlier, I described that sleep-related problems affect millions of people and are associated with billions of dollars in direct and indirect costs. What I did not mention is that social jet lag is believed to be a major root of these sleep pathologies [426]. In fact, a recent large-scale study found that more than 70% of the population suffers from significant social jet lag, with individuals’ biological and social clocks differing by more than one hour [424].

Social jet lag can additionally lead to a number of serious illnesses such as cardiovascular disease, diabetes, obesity, and cancer [266]. Shift workers, who often suffer from chronic and severe social jet lag, are more likely to experience these illnesses compared to daytime workers [385, 469]. For young adults including students, who also have an increased likelihood of circadian misalignment [426], social jet lag can additionally increase the risk of using drugs and alcohol [481, 528] and result in learning deficits [81].

Along these same lines, shift work and school schedules have been the most commonly studied culprits of social jet lag [230, 377]; but in today’s world, social demands have also begun emanating from the increasingly widespread use of digital technologies. I believe this digital connectivity may be bringing with it additional social constraints that can further disrupt our individual body clocks — an impression corroborated by some of my research that I will present in Chapter 4. Potentially, the aforementioned growing prevalence of circadian disruption may therefore be partially explained by this increasing adoption of personal devices and information technologies that implant an ethos of constant connectivity and expected availability.

Circadian disruption also has a strong association with mental health conditions and neuropsychiatric illness, including anxiety, attention-deficit hyperac-
tivity disorder, bipolar disorder, depressive disorder, obsessive-compulsive dis-
order, and schizophrenia [202]. Recent studies indicate that circadian abnormal-
ities are not only linked to a number of psychiatric disorders but that disruption
may be directly responsible for disease etiology [271, 323], for instance, poten-
tially triggering the onset of schizophrenia in susceptible individuals [238].

My work in the area of mental health focuses primarily on bipolar disorder
(BD). One of the most prominent features of BD is its rhythmicity, including
mood episodes that cycle on an approximately regular basis [464]. Circadian
instability has been identified as a contributing factor behind the development
of BD [43], and compelling evidence establishes a link between circadian distur-
bances and BD symptoms [282], including the onset of relapse after remission
[43, 171, 399]. Other mood disturbances including major depressive disorder,
seasonal affective disorder, and sundowning (a psychological condition associ-
ated with increased anxiety and agitation in patients with dementia) are also
associated with negative changes in circadian rhythm functioning [520].

3.3.4 Traditional Assessment Methods from Chronobiology

A number of methods exist to measure circadian rhythms and disruptions. Bi-
ological markers are the most accurate, but they are also the most invasive. For
example, core body temperature is considered a robust biomarker of circadian
rhythms and circadian dysregulation [116, 499], with measurement via rectal
probes being the most accurate and widely used method in the scientific liter-
ature [345]. While consistent efforts have been made to perform less invasive
assessment through wearable devices that measure oral and skin temperature
such approaches are less reliable. Hormones in the human body are also used as circadian biomarkers. Two of the most well studied are melatonin, which is measured from blood, saliva, and urine, and cortisol, which is measured in the same manner.

Next, given that sleep is both a reflector and modulator of our latent circadian rhythms, tracking sleep-wake patterns is useful in determining circadian patterns and disruptions. The gold standard for sleep monitoring is polysomnography (PSG); however, the required setup, controlled environment, and specialized equipment makes PSG infeasible for longitudinal or in-situ tracking. Instead, a wide array of studies use actigraphy, which measures body movement through the use of a wearable sensor. A number of studies have found sleep patterns inferred from actigraphy to be reliable and consistent with PSG. However, while actigraphy is less invasive than procedures associated with biomarker measurement and is more practical than PSG, it still requires a participant to wear a specialized device all day and night for the duration of the study period, which typically lasts at least 7 days but preferably spans 14 days or longer. This condition may be less problematic for laboratory or field studies of a short duration, but using actigraphy to track circadian rhythms over an extended period of time and across a large population is still difficult due to device burden and wear-compliance.

The use of biophysiological assessments such as those mentioned above are therefore mostly limited to small laboratory studies given their invasive nature. For more broad scale investigations, manual self-report via survey or diary instruments can be a more suitable approach for capturing sleep and wake patterns and the underlying circadian rhythms. One of the most prominent survey-
based instruments for assessing behavioral manifestations of circadian rhythms is the Munich Chronotype Questionnaire (MCTQ) [428]. To measure individual chronotype, the MCTQ includes questions related to sleep-wake behaviors (e.g., timing, preferences) as well as daily activities (e.g., light exposure, lifestyle details) for both work and free days. The use of the MCTQ to assess chronotype has been clinically validated in controlled settings against biomarkers and actigraphy data [427]. Another self-report method commonly used for sleep assessment is sleep logging via diaries, which a number of studies have utilized to determine sleep onset, offset, awakenings, and duration; and they are often applied as part of diagnosing and treating sleep disorders and circadian rhythm abnormalities [547]. Comparison of journaled sleep logs with actigraphy-based estimation of sleep behaviors generally shows reasonable agreement [297]; however, diarying faces limitations associated with self-report in general, including non-adherence, inconsistent completion, and potentially unreliable subjective and retrospective recall.

A number of techniques thus exist for assessing circadian rhythms and disruptions. However, the methods used in laboratory studies (e.g., that require specialized equipment or regular blood samples) are not scalable for administration to a large population. Subjective reports and surveys are more broadly deployable, but these methods are not well-suited for continuous monitoring over longitudinal periods and often fail to capture subtle details and instantaneous changes regarding the relationship between the circadian system, individual sleep patterns, and environmental effects. The ability to answer fundamental questions about sleep and circadian rhythms in real-world settings therefore depends on developing new approaches to detect and infer behavioral traits of circadian biomarkers in a low-cost, reliable, and scalable manner.
As a result, chronobiologists have pointed out the need for broad, in-situ data collection methods that can record real-time data for large populations spanning various time zones and geographical locations [423]. In my PHI work, I have attempted to answer this call by developing passive sensing approaches to assess circadian rhythms and attendant aspects of health.

3.4 Informing PHI Development

In the previous section, I reviewed the knowledge that I gathered from the domain of chronobiology. To briefly summarize, individuals have idiosyncratic circadian rhythms, which fluctuate in substantial and predictable ways. These biologically-rooted rhythms impact nearly every aspect of our functioning, from nightly sleep patterns to hour-by-hour cognitive performance to long-term physical and mental wellness. In addition, disruption to these rhythms can have a substantial negative impact on these same aspects of our overall health. Overall, this understanding of the biology behind the rhythms that guide our lives motivates PHI approaches to deeply consider this information when assessing sleep and activity. In this section, I show how I made such considerations in my chronobiology-driven case study, using gathered knowledge to determine what to assess, for whom, and how.

3.4.1 Defining Scope

As described earlier, it can be desirable for the sakes of practicality and impact to work on scoped solutions to particular sub-realms of a given health area,
rather than attempting to create a catchall PHI system. I have taken this scoped approach in my case study. As described, circadian rhythms govern nearly all aspects of our biological functioning including our sleep patterns, cognitive performance, mental health, digestion, sensitivity to pain, athletic performance, and much more [266]. Someday, I hope to see chronobiology-driven PHI systems that support all of these elements. However, my work to date has not attempted a realization of this vision in one fell swoop. Rather, I have incrementally focused on a few select facets of our chronobiology: sleep-related circadian disruption, alertness performance, and disordered mood. These areas provide breadth across physical, behavioral, cognitive, and psychological aspects of our wellness while remaining tractable.

In the context of sleep, I focused on sleep-related circadian disruption because I was motivated by the significant negative effects disruption can have on overall health, as described in Section 3.3.3. In the context of cognitive performance, where our cognitive processes are complex and multilayered, I focused on alertness because it is a primary construct in the human performance system and correlates with a number of other cognitive functions, as described in Section 3.3.1.

Next, domain knowledge about the degree of inter- as well as intra-individual differences associated with the phenomena being studied can also help in scoping. Because circadian rhythms vary from person to person and considering that chronotype can be influenced by environmental factors and changes over the course of a single person’s lifetime, different populations or even the same person over time can have drastically variable requirements from a chronobiology-driven tool. Such variability not only motivates personalized
approaches, but it also highlights the difficulty in achieving a PHI solution that adequately meets such a diverse set of needs in full — a challenge we can address by scoping and tailoring initial efforts to a specific group who stands to benefit the most from PHI technology.

For my work on assessing sleep and alertness, I made the decision to initially focus on a student population. This choice was not arbitrary or made out of convenience — rather, this group is compelling to spotlight for several reasons. Studies find these individuals suffer from chronic lost and interrupted sleep, which can lead to poorer academic performance, increased stress, and mental health problems [481]. At the same time, individuals of this age group are having to manage an increased risk for developing anxiety, depression, and other mental and emotional health problems due to academic demands and pressures to succeed that have been mounting in recent years [231].

Further, these individuals tend to be on the later end of the chronotype scale, which means they face a particularly high risk of circadian misalignment [426]. In fact, this age group experiences the most severe symptoms and consequences of social jet lag (which, as a reminder, is the work day versus free day sleep schedule instability that stems from biological sleep preferences experiencing interference from social constraints, like the early start times of traditional school schedules) [426]. As mentioned, such disruption in younger populations can increase the risk of drug and alcohol use [481, 528], produce cognitive impairments and learning deficits [81, 115], and lead to problems with attention and procrastination [130].
3.4.2 Soft Sensing for Health Assessment

Having decided what aspects of health to target and for whom, I next determined the assessment strategies I would take. In Section 3.3.4, I reviewed how chronobiologists traditionally measure circadian rhythms. This knowledge helped me decide whether I could apply these methods in a PHI system or whether I would need to devise an alternative assessment strategy. I chose the latter. Given the identified shortcomings of these traditional means of assessment and chronobiologists’ expressed need for more scalable and in-situ methods, my case study pursued a passive sensing approach to assessing sleep and alertness. More specifically, I took a soft sensing approach, leveraging technology-mediated digital traces to model individual behavior and infer these chronobiology-related aspects of health.

Generally speaking, a soft sensing approach has a number of advantages over the traditional assessment methods. First, this type of passive sensing is more affordable, as specialized and expensive equipment is not required (only the technology that individuals already possess and use). It also provides more granular data given that collection is automated, which means it can be performed continuously and over long periods of time. Further, soft sensing is less susceptible to the self-reporting biases described earlier and is also less burdensome and intrusive — although asking someone to grant access to large volumes of personal and potentially sensitive information is intrusive in another sense. To what extent this is considered intrusion is an individually-variable and open question; in my case study work, at least, people were receptive to the approach.

For data, I focused on soft sensed usage logs for smartphone apps and social media. The increasingly large and diverse user bases of these technologies were
a main draw in choosing their data. As just mentioned, a key advantage of soft sensing approaches is that they enable real-time, continuous, and longitudinal monitoring. However, this is only the case if the individual being monitored is a regular user of the technology doing the passive monitoring; otherwise data will be too sparse or skewed. The choice of these data was therefore motivated by the fact that mobile and social technologies have reached deep penetration within the target population of young adults. Specifically, the most recent statistics report that smartphone ownership has already reached 86% for U.S. “millennials” in the 18–34 year old age range, who are also the heaviest and most habituated users: 52% claim they could not last more than 24 hours without their phones; 90% sleep with or next to their phones; 54% report checking their phones “almost constantly”; and 90% check at least once an hour even during social situations such as meals, meetings, and conversations [64, 462]. Similarly, over 90% of 18–29 year olds in the U.S. hold at least one account on and regularly use social media, and that percentage is still rising [142, 281].

Beyond the broad reach of these technologies and their deep embeddedness into daily life, smartphone and social media data seemed promising for other reasons too. As I overviewed in more detail in Section 3.2.3, a growing body of research has had success in leveraging these or similar data to model traits and behaviors related to sleep [24, 217, 224, 333] as well as attention and boredom [309, 311, 383, 396, 397], which are aspects of cognition related to alertness.

I had a few additional reasons for selecting social media data in particular. One, I felt confident working with it, given favorable past experiences I had with it in the context of PHI assessment, including my work on smoking cessation described at the beginning of this chapter [355]. Second, social data seemed
appropriate for the case of assessing sleep-related circadian disruption, given that such disruptions often stem from factors that are social in nature [528]. The traditionally identified social constraints responsible for disruption are encountered in an offline context (e.g., school timings or evening social schedules), but my thinking was that technology-mediated social interactions might be a contemporary source of disruption too, as I mentioned earlier.

Further, beyond my sense that social media data was familiar and fitting, these data were also attractive because they included textual content naturally expressed in a person’s own voice (e.g., in the form of Facebook posts, messages, etc). Advances in psycholinguistics have shown the effectiveness of using the text people write to evaluate various behavioral, cognitive, and psychological attributes, with analysis tools such as the Linguistic Inquiry Word Count (LIWC) or Affective Norms for English Words (ANEW) well validated on such assessment tasks [63, 391]. Further, researchers have found strong correlations between language use and various aspects of health (e.g., with physical health [78], cognitive processes [390], and emotional wellness [436]). Linguistic analysis has also been used successfully in this way on the types of short texts typically found on social media specifically [120, 122, 265].

Altogether, this chapter has demonstrated how inquiry into domain knowledge can support decision-making during PHI development. Specifically with respect to this dissertation’s main case study, I have drawn from chronobiology to devise a plan that can be used in the framework’s next stage of health assessment — a soft sensing approach for studying sleep-related circadian disruption as well as cognitive performance in young adults. The next chapter, “Domain-Driven Health Assessment”, presents that assessment work.
In the previous chapter, I overviewed a process of domain inquiry: selecting a compelling area for applying PHI technology, gathering salient domain knowledge, and using that knowledge to inform subsequent stages of PHI work, including a data collection and analysis strategy. In this chapter, that analytic plan is executed. (See “Domain-Driven Health Assessment” portion of Figure 2.1).

To demonstrate this stage of the framework in action, I continue to use my case study on developing chronobiology-driven PHI. Specifically, in this chapter I present two experiments undertaken to investigate how technology-mediated digital traces can be leveraged (i.e., passively collected and analyzed) in order to assess idiosyncratic biological rhythms. In the sections that follow, I first overview the protocols of these experiments, participant characteristics, and data collection procedures. I then present findings from each experiment, along with chronobiology-guided interpretations of their results. I conclude this chapter with a discussion of these experiments, their limitations and future opportunities, and broader implications for the domain-driven framework.

4.1 Method

In Section 3.4.2, I explained the merits of using passively sensed data from smartphones and social media in the context of assessment. In Section 3.4.1, I also justified a decision to scope collection and analysis of these data to a population of young individuals, who could particularly benefit from chronobiology-
aware PHIs and for whom this sensing methodology is particularly apt. In the following subsections, I overview how these and other relevant data were collected from samples of this target group recruited to participate in two experiments, which I refer to as Experiment 1 [352] and Experiment 2 [353].

Both experiments examined the interplay between biological rhythms and technology use. Specifically, Experiment 1 explored how social-sensor data can be leveraged to detect sleep-related behaviors and circadian disruptions, and it took preliminary steps toward analyzing the impact of inadequate sleep on cognition and mood. Digging deeper into daily functioning, Experiment 2 built on chronobiology about cognitive performance rhythms in order to explore and interpret a number of relationships among smartphone use, alertness, sleep, and latent biological traits. Findings from both experiments contribute to chronobiology-driven assessment by identifying ways to capture and analyze usage patterns in order to passively monitor idiosyncratic biological rhythms.

4.1.1 Participants & Procedures

For both Experiment 1 and Experiment 2, public mailing lists, recruitment portals, and snowball sampling were used to recruit participants who were in the target age group, had been using smartphones for at least six months prior to the beginning of the experiments, and were willing to participate for the experiments’ full durations.

Experiment 1’s sample consisted of 9 participants (7 males, 2 females) aged 19–25 years old, and the study lasted 97 days from November 22, 2013 – February 26, 2014. Given that this experiment was interested in exploring how social
interactions and socially-defined demands impact circadian patterns, this study spanned three phases of student life with varying scheduling constraints and social environments: end of Fall semester (34 days), Winter break (24 days), and start of Spring semester (39 days). All participants had standard class schedules, except for one person who had an internship and attended no classes during the Fall semester. Experiment 2’s sample consisted of 20 participants (7 males, 13 females) aged 18–29 years old, and the study lasted 40 days from February 13 – March 24, 2015.

To onboard participants, they were invited to the lab, where I or a colleague explained procedures and installed, tested, and demonstrated the experiments’ data collection tools on their phones. At the end of the experiments, participants were compensated based on the number of completed sleep diary entries, the amount of successfully logged data, and the number of conducted interviews. (I describe all these data further in the following subsection). All collected data were anonymized and encrypted, and Cornell’s Institutional Review Board approved all procedures for both experiments.

4.1.2 Data

Survey Measures

Chronotype. As previously described, an individual’s chronotype reflects his or her unique circadian rhythms, which underlie numerous biological and behavioral processes including sleep and daily performance. To measure chronotype, participants took the Munich Chronotype Questionnaire (MCTQ) [428] during recruitment. The MCTQ includes questions about sleep-wake behaviors
(e.g., timing, preferences) as well as daily activities (e.g., light exposure, lifestyle
details) for both work days (days on which alarms are used) and free days (days
without an externally-imposed work or school schedule — typically weekends)
[424]. The use of the MCTQ to assess chronotype has been clinically validated
in controlled settings against biomarkers, actigraphy data, and sleep logs [427].

To provide a quantified, comparable representation of chronotype, the
MCTQ estimates chronotype based on the halfway point between sleep onset
and waking on free days [528] ($MS_F$). Previous studies have found this mid-
sleep point to be the best phase anchor for biochemical indicators of chronotype
[483]. $MS_F$ is corrected ($MS_{F_{SC}}$) to account for longer sleep durations taken on
free days; that is, except for extreme early chronotypes, most people accumulate
sleep debt during work days and then compensate (if possible) by oversleeping
on free days [428]. Thus, chronotype is a continuous variable quantified as:

$$MS_{F_{sc}} = MS_F - 0.5 (SD_F - (5 * SD_W + 2 * SD_F)/7)$$

where $SD_F$ and $SD_W$ are sleep duration on free days and work days, respec-
tively, and $(5 * SD_W + 2 * SD_F)/7$ provides average sleep duration over a week.

Figure 4.1 shows chronotype according to $MS_{F_{SC}}$ for each participant in Ex-
periment 1, and Figure 4.2 illustrates the distribution of chronotypes for partic-
ipants in Experiment 2. Both figures’ early-late key is based on an established
early-late spectrum for a general population [426].

For such a general population, average $MS_{F_{SC}}$ falls closer to the 4:00–6:00
range. Here, participants trended later (average $MS_{F_{SC}} = 5:34$ in Experiment
1 and average $MS_{F_{SC}} = 5:56$ in Experiment 2). This was expected given their
ages; and considering the narrow chronotype range typically associated with
this age group, the samples actually provided a relatively wide variability of
chronotypes [426, 428]. To verify that these samples were representative of my population of interest, I also administered the MCTQ to over 200 additional students for comparison. Generating an age and sex matched random sample from that N=281 large survey gave a mean $MS_{SC} = 05:46$, which is indeed similar to the average $MS_{SC}$ of participants in both Experiment 1 and 2.

Figure 4.1: Chronotypes of participants in Experiment 1.

Figure 4.2: Distribution of participant chronotypes in Experiment 2.
In determining a fair “early” versus “late” threshold for a particular population of interest, chronobiologists consider the chronotype distribution of that group [425]. Thus I followed prior work [422] and treated $MSF_{SC} \leq 5:00$ as early and $MSF_{SC} > 5:00$ as late, given participants’ young age range and corresponding later-skewed chronotype distribution [203]. This split created groups acceptably balanced in size and also provided the highest level of agreement among MCTQ-measured chronotype, self-perceived lateness reported during interviews, and earliness/lateness assessed via the Morningness-Eveningness Questionnaire [213], which participants also completed during recruitment as an additional check on their early/late classifications.

**Personality.** Given that previous research has found associations between Big Five personality dimensions [111] and individual differences in circadian rhythms [6], personality was captured as a control variable using the Big Five Inventory (BFI) [226]. All personality factors showed good internal consistency within participants (Cronbach Alpha between 0.74 – 0.89), and correlations between participants’ chronotype and personality were similar to findings from prior studies [6, 511].

**Daily Self-Reports**

**Sleep Logs.** Throughout both experiments, each participant maintained a once-daily online sleep diary. Guided by sleep diaries from prior work [127], questions asked about the previous night’s bedtime, number of minutes to fall asleep, number of wakeups during the night, details about any experienced sleep disruptions, wake time, perceived feelings upon waking, presence and duration of groggy feelings after waking, and overall alertness and sleepiness.
Participants received a reminder each morning to record this information; reminders were sent via an email in Experiment 1 and via a mobile notification in Experiment 2. Compliance rates were 76% and 73% in Experiments 1 and 2, respectively. To ensure data quality, any retrospective entries were discarded and only those that recorded the previous day’s sleep were retained. Though tedious to collect, self-report journaling has been validated by prior studies as a reliable approach for in-situ measurement of per-night sleep [58, 333, 389], and is considered less intrusive than body-based sensors such as actigraphy.

Alertness EMA. Since Experiment 2 was interested in more deeply investigating circadian rhythms related to daily performance (specifically with respect to relationships among alertness, sleep, and smartphone use), participants in this study also completed a brief ecological momentary assessment (EMA) of alertness on their phones. Specifically, the study used a three minute version of PVT-Touch [243], a validated smartphone-based psychomotor vigilance task (PVT) for objectively measuring alertness. The PVT, which is sensitive to changes in alertness [35] and is immune to practice or learning effects [270], measures alertness by displaying a visual stimulus and recording the elapsed milliseconds before a tactile response.

The alertness EMA was delivered four times daily at the start of time windows defined by prior work [1] for morning, afternoon, evening, and late night to increase the breadth of coverage across the day. Participants could complete the assessment anytime within the time window, providing further variation in the collection times. For the morning, afternoon, and evening windows, the average compliance rate was over 75%; the late night window overlaps with sleep [230] and thus had an expected lower coverage of 14%.
Following an established formula from the chronobiology literature, I computed alertness performance for a given session according to its percent deviation from that individual’s baseline, where individual baseline was computed as the mean reaction time across all test sessions, after removing false starts and outliers 2.5 standard deviations above or below the mean [504].

Usage Logs

A core component of both experiments was technology usage data.

*Computer-Mediated Communication (CMC) Data.* In Experiment 1, participants installed a smartphone application a colleague developed to run in the background and collect usage data. My analysis of this data focused on the probes for technology-mediated social interactions: phone calls, text messages, and social media app usage. For Experiment 1, I also requested participants’ permission to download their Facebook friend data along with their logs of status updates, posted comments and “Likes”, “Ask a Question” posts, location check-ins, and outgoing private messages. I refer to all these Facebook data as “posts”. Interested in participants’ Facebook interactions with other users, I filtered out system-generated posts (e.g., tagged photo alerts). I focused on Facebook since it was the most popular social network used by all of the study’s participants, but it could be desirable in future work to incorporate data from additional social media platforms to further verify and compare results.

*Smartphone Application Use Data.* In Experiment 2, participants also installed the robust AWARE framework, which provides fine-grained sensing of mobile application usage [159]. I captured the timestamped log of any app coming to the foreground (e.g., app launches or switches), which I refer to as “usage
events”, along with the duration of use. Most of my analyses focused on the instances of app foregrounding since prior work shows this is an informative portrayal of usage behaviors [54, 66, 199, 228, 454] and is less prone to measurement errors than metrics like duration [134]. This also facilitated comparison with related research using the same metric. As they are not indicative of user behavior [54], I disregarded background apps with which the user did not interact as well as system-generated activity such as automated notifications.

To categorize participants’ logged applications, I followed prior research [54] and used the app’s developer-specified category in the official Android application market, Google Play, where each app was associated with a single category.

I filtered out the Tools category since its apps related to launcher processes, system activities, or settings, which were either not user-originated actions or did not provide the sorts of insight I desired into individuals’ self-driven usage behaviors [54], with the exception of clock and weather apps, which I relabeled into a new Time & Weather category. I also filtered out Health & Fitness apps since they comprised less than 1% of usage events and were used by a minority of participants. Future work could do well to specifically recruit Health & Fitness app users in order to explore the relation between app usage and physical performance, which also exhibits well-known circadian fluctuations [140].

Again following prior work [54], I separated web browsers and email apps from Communication into more fine-grained Browser and Email categories. Then, to further facilitate analyses and because I hypothesized that circadian rhythms of cognitive performance might be most strongly reflected by productivity and entertainment-oriented application use, I consolidated a number of related apps into higher level categories: Entertainment contains apps originally categorized
<table>
<thead>
<tr>
<th>Category</th>
<th>Example Apps</th>
<th># of Apps</th>
<th># of Usage Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browser</td>
<td>Chrome, Firefox</td>
<td>10</td>
<td>17683</td>
</tr>
<tr>
<td>Communication</td>
<td>Facebook Messenger, GroupMe, Phone, SMS</td>
<td>33</td>
<td>32906</td>
</tr>
<tr>
<td>Email</td>
<td>Gmail, Inbox</td>
<td>3</td>
<td>5142</td>
</tr>
<tr>
<td>Entertainment</td>
<td>Clash of Clans, Ebay, Netflix, YouTube</td>
<td>60</td>
<td>9863</td>
</tr>
<tr>
<td>Productivity</td>
<td>Evernote, OfficeSuite, To Do Reminder, Piazza</td>
<td>47</td>
<td>3146</td>
</tr>
<tr>
<td>Social Media</td>
<td>Facebook, Twitter, Yik Yak</td>
<td>14</td>
<td>27693</td>
</tr>
<tr>
<td>Time &amp; Weather</td>
<td>Clock, Timely, Weather Channel</td>
<td>12</td>
<td>1702</td>
</tr>
</tbody>
</table>

Table 4.1: Categories of applications used by participants in Experiment 2, with examples and amounts of applications and usage events.

as Entertainment, Games, Media & Video, Music & Audio, Photography, or Shopping; and Productivity contains apps originally categorized as Productivity, Business, Education, or Finance.

Manually inspecting all apps within each category, a colleague and I independently verified, discussed, and came to full agreement that similar kinds of apps were folded together and that each resultant parent category fairly represented its contained applications. These categories are shown in Table 4.1, along with information about the unique number of apps participants used from each category and the total number of usage events.

**Interviews**

Both experiments included interviews with participants. In Experiment 1, I or a colleague interviewed each participant three times: an initial interview upon recruitment, a second interview at the end of the Fall semester prior to the start of Winter break, and a concluding interview at the end of the study. Participants were asked to discuss various aspects of their phone use, especially
in the morning and at night and across various contexts, as well as motivations and any perceived patterns to their social media use specifically. In Experiment 2, I or a colleague interviewed each participant upon recruitment and again upon completion of the study. Questions asked about sleep-wake behaviors and perceived connections among one’s alertness, fatigue, sleep, and time of day. We also asked about technology usage habits, including thoughts about technology’s impact on alertness, fatigue, or sleep. Finally, we asked about experiences with productivity software and reactions to ideas for chronobiology-aware tools.

These interviews provided the opportunity to verify assumptions, validate analyses against a self-reported “ground truth”, and seek explanations behind participants’ observed technology-mediated behaviors. Interview data were qualitatively analyzed using thematic analysis [61].

In the following sections, I report on the findings from both experiments, contextualizing quantitative results with insights and representative quotes from interviews where appropriate.
4.2 Experiment 1 Findings

4.2.1 Daily Rhythms in Sleep and Social Technology Use

To begin, I analyzed data from phone probes and social media logs (together referred to as “CMC”) to gain a sense of participants’ typical usage trends. I also compared my observations to results from prior studies (consistently finding close alignment) in order to further support that the experiment’s small-scale sample was representative of college students more generally.

Specifically, I observed the daily usage trends shown in Figure 4.3. Usage was heaviest in the late evening, until about 11pm. Levels of social media app use and Facebook posting activity in particular continued slightly later until around 1am. These observations align well with prior studies on CMC use, which find that Facebook usage increases through the evening until around midnight [310], that social mobile applications have the highest probability of being used from 9pm to 1am [54], and that text messaging frequently occurs late at night and causes later bedtimes [495]. Following this CMC use, sleep diary

Figure 4.3: Daily trends in participants’ average CMC-based usage.
entries indicated participants go to sleep within an average of 49 minutes; prior research similarly finds sleep occurs within 60 minutes of computer use for 60% of 19–29 year olds [364].

Individuals in this 19–29 age range are known to go to sleep later than any other age group, and adolescents in particular tend to delay their bed and wake times as well as suffer from decreased sleep length and increased sleep irregularity [113, 310]. Indeed, the average sleep onset of participants in Experiment 1 ranged from 1:36–2:14am (depending on weekday or weekend, semester or vacation) — quite late timings, likely since they are on the younger side of this age range. Additionally, I found less than half (49.1%) of reported sleep durations to be 7 or 8 hours, and 23.3% of reported durations were 6 hours or less. These findings are close to those observed in prior studies [347] and indicate a concerningly high incidence of insufficient sleep among participants. I also found 15.2% of sleep durations to be 10 hours or more, which is further troubling given that exceedingly long as well as short sleep durations are detrimental to physical and mental health and are associated with a range of problems related to academic performance, reckless behavior, and substance abuse [209].

Researchers have suggested that such sleep inadequacy may in part be due to increased usage of the internet and social media [249]. Experiment 1 found similar results — that social media may not only reflect but also modulate delayed sleep onsets. Specifically, on nights when participants used social media apps and posted to Facebook after 12am, they reported an average of 34 minutes less sleep. Using participants’ sleep diaries, I also compared each night’s number of reported sleep interruptions to the timing and amount of social media use the prior day. I found that for nights during which participants experienced one
or more sleep interruptions, they used social media nearly twice as much the day before (1.8 times more on average; Wilcoxon sign-rank test, $p < .001$). Diary entries suggested that such behavior produced feelings of tiredness, as my analysis showed that late-night social media use was associated with reports of feeling “fatigued” as opposed to “refreshed” ($\chi^2 = 10.21, df = 1, p < .05$). Facebook updates from the following day sometimes expressed similar exhaustion — as three examples: “Super sleepy”, “I am exhausted”, and “So tired and really want another hour to sleep”.

During interviews, many participants explained that CMC use had become a routine part of bedtime habits. For example, “I do my before-sleep routine, get into bed with my phone, spend about fifteen minutes on Facebook, then set my alarm, put the phone under my pillow, and am asleep”. The exact ways in which technology was incorporated into late night behaviors varied across individuals depending on lifestyle aspects or chronotype traits. For instance, all participants in relationships reported using CMC to communicate with partners just before bed. Late types noted using social media as something to do when unable to fall asleep (due to their late biological clock), while as expected the early type participant disagreed, “People usually keep me up not technology”, referring to evening social schedules, which are shaped more by late types [426]. Most participants also expressed that social media keeps them up longer than planned, for common reasons such as “endless scrolling” social feeds that make them “feel like an addict, obligated” to “need to know what’s going on”.

Based on these findings that CMC use related to sleep characteristics such as length and quality, I next explored leveraging this data for sleep sensing.
4.2.2 Leveraging Social Data for Sleep Sensing

Inferring Sleep Onset, Duration, and Waking

I first attempted to infer sleep events from CMC patterns by implementing the sleep-inference algorithm built on screen on/off patterns presented in prior work [3]. I instantiated the algorithm using participants’ phone probe data, social media app use logs, and Facebook posts in order to model sleep events according to the longest nightly gaps in usage. I pre-processed these social-sensor inputs to filter usage events before 10pm or after 7am, which do not normally coincide with sleep periods for non-shift workers [230]. I also eliminated any usage events with a duration of less than 30 seconds, which are likely due to automated phone notifications rather than active user interactions [3].

Table 4.2 presents the accuracy of this sleep duration inference compared with the screen on/off approach and with participants’ ground truth sleep diary data.

<table>
<thead>
<tr>
<th></th>
<th>Social Data</th>
<th>Screen On/Off</th>
<th>Ground Truth Diary</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>8.44*</td>
<td>8.54*</td>
<td>8.13</td>
</tr>
<tr>
<td>P2</td>
<td>7.64*</td>
<td>8.09</td>
<td>7.45</td>
</tr>
<tr>
<td>P3</td>
<td>8.21*</td>
<td>8.33*</td>
<td>8.15</td>
</tr>
<tr>
<td>P4</td>
<td>7.53*</td>
<td>8.02*</td>
<td>7.25</td>
</tr>
<tr>
<td>P5</td>
<td>6.11*</td>
<td>5.44*</td>
<td>6.12</td>
</tr>
<tr>
<td>P6</td>
<td>7.15*</td>
<td>7.17*</td>
<td>7.13</td>
</tr>
<tr>
<td>P7</td>
<td>7.63</td>
<td>7.16*</td>
<td>7.14</td>
</tr>
<tr>
<td>P8</td>
<td>7.38*</td>
<td>7.30*</td>
<td>8.14</td>
</tr>
<tr>
<td>P9</td>
<td>7.48</td>
<td>5.42</td>
<td>6.25</td>
</tr>
</tbody>
</table>

Table 4.2: Average sleep duration for each participant according to social-sensor data, screen on/off data, and ground truth sleep diary data. (* denotes inferences that fall within 95% confidence interval based on diary self-reports, p < .01).
aries. Results show my technique’s reliability, which achieved an average difference of only 23 minutes between socially-sensed and self-reported sleep duration. This prediction was more accurate than from screen on/off alone [3]; and it also managed to outperform more complex algorithms based on environmental factors such as light, movement, and sound as well as phone locking and charging events [89]. My approach is thus desirable for a few key reasons. First, this technique is as reliable yet more unobtrusive and computationally lightweight than those built upon frequent momentary assessments (EMAs), heavy instrumentation, or the use of wearable sensors. In addition, by leveraging web data, it is able (unlike approaches based solely on mobile sensor data) to continue capturing signals about a user’s behavior even if she is interacting through a device other than her personal phone such as a tablet, desktop computer, friend’s device, or public computer.

My approach did overestimate sleep when the stop and start of CMC use did not precisely adjoin sleep onset and wake, respectively. By incorporating an error term to the calculation of sleep duration per participant (based on chronotype and individual differences in pre-bed and post-wakeup CMC usage learned from the study’s first week of data), I was somewhat able to correct for this non-usage gap, and more complex learning could further improve accuracy. Conversely, there was sometimes an underestimation in sleep duration when notifications were mistaken as active usage. By incorporating a threshold for minimum usage duration, I attempted to filter out such device-generated events, but more sophisticated instrumentation could further help eliminate such misinterpretation of phone events that are not indicative of genuine user activity.
Interview data allowed me to uncover other points of failure and opportunities for improvement. For instance, one participant described pre-bed phone use as a common tendency, identifying watching movies and using Twitter as typical nightly activities; and she also noted normally checking email and texts upon waking. Similarly, another participant explained that morning phone use involved weather and calendar checking, and he discussed playing video games before bed but explained that he did so on a desktop computer rather than the phone. Thus incorporating into sensing both additional forms of social data (e.g., Twitter, email) as well as broader non-CMC usage data (e.g., app logs, web histories) and from across additional devices would be straightforward next steps toward better sleep-event estimations.

This ability to infer sleep onset and waking introduces opportunities to assess a number of other chronobiology variables that can be derived from this information, such as chronotype, which is quantified using sleep duration and midpoint (both of which can be computed from onset and waking). Given the scope of this experiment, I focused on using this sleep sensing technique to assess circadian disruptions related to social jet lag (defined in Section 3.3.3).

Assessing Social Jet Lag

As I discussed earlier, social constraints can result in later sleep onsets and earlier required wake times that are in opposition to our own internal timings.

Figure 4.4 shows the average social-sensed sleep duration on work days and free days for each participant — and illustrates the discrepancy between the two. Duration is calculated as the amount of time between sleep onset and waking [377]. (Note that participants’ work days were Monday through Friday,
and their free days corresponded to Saturday and Sunday; but generally speaking, work and free days do not necessarily have to coincide with the standard workweek and weekend days).

Chronobiologists refer to this reversed sleep pattern for early and late chronotypes on work and free days as the “scissors of sleep” [422] — a phenomenon my sensing approach was able to reveal. Specifically, these results demonstrate that for later chronotypes in the sample (e.g., P1, P2, P5), sleep duration was systematically shortened on work days, which led to accumulated sleep debt that was then compensated for by sleeping more on the weekend. This same effect has been observed in past research [428]. Excluding the early chronotype participant (P4), participants slept an average of 67.8 minutes more on weekends. In contrast, P4 exhibited precisely the opposite pattern. For this
individual, longer durations of sleep happened on weekdays while sleep was shortened on the weekend. This was likely because P4’s work week schedule fit better with his internal timing preferences while his weekend sleep was forced to shift due to social engagements with later-type peers. Indeed, sleep onset for P4 was 93 minutes later on weekends than during the week, plus sleep duration was reduced (by an average of 54 minutes) since the natural circadian drive would still prompt an early wake up even after a later-than-preferred sleep onset [426] following a night of socializing.

Next, to quantify social jet lag and assess its severity across participants, I computed the difference between mid-sleep (the halfway point between sleep onset and waking) on free days (MS\textsubscript{F}) and on work days (MS\textsubscript{W}) per a formula from the chronobiology literature [528]:

$$\Delta MS = |MS\textsubscript{F} - MS\textsubscript{W}|$$

Figure 4.5 shows the results of this calculation using the social-sensor data, presented according to participant chronotype. Alarmingly, it is estimated that

![Figure 4.5: Socially-sensed average social jet lag (discrepancy between mid-sleep on free days and work days) across chronotypes.](image-url)
over 70% of the population suffers from social jet lag [424], and I unfortunately observed it impacting Experiment 1’s participants too. My results compare well to those from prior analyses on the MCTQ database, which similarly find social jet lag ranging from approximately 1–2 hours [425]. In addition, I found that the extreme ends of the chronotype spectrum experienced more social jet lag; and it was most severe for the sample’s later types, as expected since their socially-constrained days (work days) outnumber their free days (weekends) [528].

I also compared social jet lag across the three phases of Experiment 1 (Fall semester, Winter break, and Spring semester) since academic responsibilities, employment schedules, and social expectations vary across these periods. Figure 4.6 illustrates results. During the Fall and Spring semesters, sleep midpoint was much earlier on weekdays versus weekends since imposed class schedules forced earlier wake up times during the week. Further, it appeared more sleep debt accumulated during work days in the Fall compared to the Spring semester, as reflected by a considerable shift in weekend sleep midpoint during Fall weekends in order to compensate. I believe this was due to the fact that the Fall study phase overlapped with the demanding end-of-semester exam period.

![Figure 4.6: Shifts in sleep midpoint across study phases.](image-url)
whereas the Spring study phase was during the (slightly) less intensive start of the semester. A number of Facebook posts from the dataset suggested this to be the case as well, for example: “lab exam. how much should i stay up to study tonight??” (Fall) compared to “still just shopping for classes” (Spring).

On the other hand, when these external academic pressures somewhat subsided during the Winter break and participants could more freely choose their sleep timings, I found far less fluctuation between sleep midpoint on weekdays versus weekends, which differed then by less than 10 minutes. Still, it is possible for individual differences to exist in terms of social dynamics during vacation periods. From interviews, I learned that some participants’ main social groups were located in the place to where they were travelling, resulting in a substantially reduced need to use CMC technology to stay in touch, as compared to while away at school and disconnected from those groups. On the other hand, other participants explained that their online networks were mainly comprised of schoolmates, which meant that leaving during the break instead resulted in increased CMC usage to maintain contact. Such results serve as an important reminder to avoid generic assumptions regarding technology usage and instead to consider the variety of individual circumstances that can impact or reflect widely different usage habits.

Nightly Sleep Disruption and Morning Sleep Inertia

A number of potential factors can contribute to sleep disruption. As mentioned, caffeine, exercise, napping, and alcohol are commonly studied by HCI researchers, sometimes with an eye to designing technology to help users maintain sleep hygiene. Regardless of its culprit, the detrimental physical, cognitive,
and psychological effects of poor sleep are numerous [81, 377]; and such deficiencies that follow a night of inadequate sleep can be initially observed during the wake up process. Specifically, the term “sleep inertia” is used to describe the time a person takes to become fully awake and functional, and prolonged sleep inertia is a symptom of social jet lag [428]. Given that the duration of morning technology usage has been shown to be a reasonable proxy of sleep inertia [3], I investigated what specific technology-mediated activities typically comprised morning usage for participants, along with the feasibility of using social-sensor data to model this sleep-wake transition.

Analyzing rise time usage, I found that all participants used their smartphones within 10 minutes after waking up for activities such as browsing the internet, checking email, and interacting with social media or communication apps. Note that this usage was separate from alarm-related usage (7 of 9 participants in Experiment 1 reported using their smartphones as their daily alarms). The amount of morning phone use I observed is consistent with prior large-scale studies on college students’ mobile device habits [279]. Prior work has also found that communication applications are typically among the first apps used upon waking from sleep [54], a tendency I found in Experiment 1’s phone probe and social media data as well: on average, I detected some form of technology-mediated social interaction within an hour of waking, with text messaging being the predominant form of social technology use (compared with phone calls and social media) on more than two-thirds of mornings.

I attempted to operationalize sleep inertia according to the duration of morning CMC activity but did not observe the same strong association found in prior work that bases usage on screen on/off events [3]. This suggests CMC-based ac-
Activities are a viable option for assessing wake events since they are frequently a user’s first form of usage upon waking — but that attention soon turns to other sorts of usage that may be more informative for measuring sleep inertia. For instance, interviews revealed such usage often involved browsing news, weather, and videos for this study’s participants. Going forward, it would therefore be worthwhile to capture data about these types of interactions if the goal were to build models for predicting morning inertia and transitional states out of sleep.

4.2.3 Sleep’s Links with Neurobehavioral Functioning

As described previously, social jet lag has numerous detrimental consequences, with symptoms manifesting as cognitive difficulties and emotional problems. Moving beyond morning rhythms, I therefore next used CMC data to explore the impacts of sleep on such neurobehavioral functioning the following day, specifically focusing on attention, cognitive functioning, and mood. These characteristics are known to exhibit strong circadian patterns, suffer substantially after sleep loss and interruption, and are considered especially important attributes to evaluate for individuals in the participants’ age group [377].

I defined a number of socially-sensed variables in order to operationalize activity levels, social interactions, cognition, and emotions, all of which prior research and this own study suggested as relevant to performing such circadian assessments. Here I present my analyses that revealed meaningful differences in these variables on days following nights of varying sleep quality. Comparisons were performed on medians using Wilcoxon sign-rank tests. Following established guidelines, I treated sleep durations lasting 7–9 hours as “adequate” and
durations outside this range as “inadequate” [88] — though it is important to note that just as our internal biological clocks direct our preferred sleep timings, there are individual differences in sleep need as well [428].

Attention and Cyberloafing

A strong theme that emerged from the interview data was that participants often turned to CMC when tired, bored, or unable to pay attention to tasks at hand and that CMC was frequently used as a way for participants to procrastinate, entertain themselves, or simply pass the time. Cyberloafing is a term used to refer to such procrastination and idling behaviors [289]. This tendency to postpone tasks may be explained by a lack of attention and an inability to focus that stem from insufficient self-regulatory resources, which drain over the course of a day and require adequate sleep to become restored [37]. Both sleep quantity and quality are important to this restoration [218], and an individual’s failure to obtain both can result in increased levels of cyberloafing [351].

To represent cyberloafing behaviors, I therefore computed the following measures based on a participant’s technology-mediated social interactions, which I refer to as “CMC usage events”:

- **Volume:** The total number of CMC usage events a participant performs in a given day between initially waking and eventually going to sleep.
- **Burstiness:** The maximum number of CMC usage events a participant performs in any single hour between wake and sleep.
- **Frequency:** The number of hours between a participant’s successive CMC usage events.
As presented in Table 4.3, my analysis of these variables found that inadequate levels of sleep were associated with heavier use of technology the following day. Specifically, nights of insufficient sleep were associated with more CMC-based usage events the next day, which were made more frequently and in tighter temporal bursts. Correlating hours of sleep with the amount of next-day cyberloafing activity showed the same negative relationship ($r = -.52, p < .01$).

During interviews, participants all mentioned checking social media when having trouble focusing or concentrating, which they expressed often happens when tired (e.g., “If I’m more tired, I’m less able to pay attention in class and more likely to use phone to avoid falling asleep or get bored more easily.”)

Prior research has indicated that individuals with higher levels of conscientiousness may naturally possess more self-regulatory resources [111] and be less susceptible to cyberloafing following lost or disrupted sleep. I therefore performed linear regression between sleep duration and the amount of subsequent CMC activity while controlling for personality. I found sleep duration ($\beta = -.39, p < .001$) and conscientiousness ($\beta = -.16, p < .01$) to be significant predictors of subsequent CMC usage; and the negative direction of the partial slopes again indicated that the less sleep an individual got, the more she used CMC technologies the following day.

<table>
<thead>
<tr>
<th></th>
<th>Adequate</th>
<th>Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume **</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>Burstiness ***</td>
<td>6.12</td>
<td>9.54</td>
</tr>
<tr>
<td>Frequency ***</td>
<td>0.71</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Table 4.3: Median values of CMC-based activity levels following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name (\(**p < .001, ***p < .0001\)).
Cognitive Functioning

Sleep deprivation can also impair cognitive performance, including in academic settings, which is particularly salient for the students that comprised this experiment’s sample. While people’s performance levels naturally fluctuate throughout the day, sleep loss dampens overall performance; and the impairment effects of fatigue coupled with these endogenous changes in daily brain function have even been equated to alcohol intoxication [269]. Conversely, adequate sleep duration improves learning and problem solving [507].

As a proxy for cognitive functioning, I assessed text-based content from participants’ Facebook posts. I first performed standard pre-processing on these posts (e.g., removing punctuation and URLs, handling spelling errors) and then calculated the following measures, which represent the sophistication of a participant’s posts and the cognitive effort required by the writing:

- LIX: A measure that indicates the linguistic sophistication of a piece of text, computed as the percentage of words having 7 or more letters plus the average number of words per post [47].
- TReDIX: A LIX-based measure adapted for use with social media content, computed as a ratio of the total count of words having 7 or more letters that appear in all posts made within a time period over the total number of posts made in that time period [220].

As summarized in Table 4.4, I found that an adequate number of hours of sleep related to higher levels of complex thought according to both cognitive functioning measures. The greatest difference was in the TReDIX measure, and linear regression confirmed a positive relationship; that is, the fewer hours of
<table>
<thead>
<tr>
<th></th>
<th>Adequate</th>
<th>Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIX *</td>
<td>0.3592</td>
<td>0.3003</td>
</tr>
<tr>
<td>TRedIX **</td>
<td>0.2738</td>
<td>0.2144</td>
</tr>
</tbody>
</table>

Table 4.4: Median values of cognitive functioning variables following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name (*$p < .05$, **$p < .01$).

sleep, the lower the subsequently demonstrated cognitive ability according to social-sensor based assessment ($\beta = 2.17$, $r^2 = 0.12$, $p < .001$).

**Mood**

Consequences of sleep reduction include negative mood, tension, nervousness, and irritability [377]. Conversely, extending sleep improves mood [232]. To evaluate whether social-sensor data could be used to reflect patterns in mood, I again turned to Facebook post data and this time applied psycholinguistic analysis techniques to compute the following measure, which prior work has found to be a reliable representation of sentiment based on its strong correlation with sentiment ratings from human judges [220]:

- Sentiment Intensity Rate: A measure of how intensely positive or negative emotions are, computed as the ratio of the sum of valence intensity of positive or negative language in posts to the total number of posts in a period. Valence intensity can be determined from the ANEW dictionary [63] and positivity and negativity from the LIWC dictionary [390].

To avoid skewed results due to participants with many more Facebook posts than others, I normalized values of sentiment variables to between 0 and 1 (scaling in this way also makes results more interpretable — values closer to 1
Table 4.5: Median values of sentiment expressed in Facebook posts following nights of Adequate vs. Inadequate sleep. Significant differences in medians marked on variable name (**p < .0001).

<table>
<thead>
<tr>
<th></th>
<th>Adequate</th>
<th>Inadequate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Sentiment Intensity ***</td>
<td>0.5373</td>
<td>0.3057</td>
</tr>
<tr>
<td>Negative Sentiment Intensity ***</td>
<td>0.4176</td>
<td>0.8388</td>
</tr>
</tbody>
</table>

indicate levels of the sentiment variable are nearer to the maximum value ever observed for that individual and values closer to 0 indicate levels nearer the minimum). Table 4.5 shows the differences in positive and negative sentiment expressed after adequate and inadequate sleep.

I found that positive sentiment following nights with adequate sleep was 1.75 times higher than following nights with inadequate sleep, after which negative sentiment was instead over twice as high. Figure 4.7 illustrates the difference in negative sentiment on days following nights of varying sleep duration. A similar relationship between insufficient sleep and negative affect has been observed in prior studies that required participants to take daily EMA-based mood assessments [347]. Interviews agreed, with Experiment 1’s participants consistently noting their usage was higher when energy and mood were lower.

![Figure 4.7: Sleep duration and sentiment the following day.](image-url)
(e.g., feeling “more down” or “down and frustrated”) and also describing using social media to “vent” or seek social support when tired and irritated.

Subsequent daily mood also showed an association with the timing of sleep onset, with 3am showing the strongest signal during exploratory analysis. Specifically, I found that participants whose final CMC activity happened after 3am had the lowest levels of measured sentiment the following day, while posts from individuals who went to sleep at a relatively earlier time were 2.2 times more positive the following day. (Note that “earlier” here means relative to other participants’ timing of sleep onset and does not necessarily mean “early” as in “early to bed and early to rise”, given these participants have later chronotypes compared to a general population). More obtrusive studies administering end-of-day mood surveys and employing a wide array of sensors (e.g., computer logging, heart-rate monitors) have similarly found that people who go to bed earlier are also happier [310], and my observation also aligns with prior work associating late-night social media usage with depression and stress [121, 355] — though the cause versus effect remains unknown. My results thus complement prior findings about a connection between sleep and mood as well as demonstrate how social media data can reveal this relationship.
4.3 Experiment 2 Findings

4.3.1 Smartphone Application Usage Patterns

Similar to Experiment 1, I began my analysis by exploring the types of smartphone apps participants in Experiment 2 used, along with temporal trends in those usage patterns both within and over days. When possible, I again compared my findings to those from prior work, both to affirm the sample was representative of larger populations and to highlight and interpret new findings.

Daily Rhythms in Smartphone Application Usage

Aggregating participants’ smartphone app use events, I found the trends illustrated in Figure 4.8, with app use overall at its lowest in early morning, steadily rising and remaining relatively high from approximately noon until late evening, and then dropping off. These trends are similar to those observed in prior studies on daily mobile, computer, and internet usage [54, 310].

The most heavily used types of applications across all hours of the day were communication and social media apps. Communication app use was highest between late morning and midnight, with peaks mid-afternoon and evening, similar to trends Experiment 1 found for phoning and texting. Usage of social media apps (which, compared to those in the communication category, are used more for consuming social content rather than communicating) had maximum usage levels between 7pm and midnight, similar to findings observed in prior studies [54, 310] as well as Experiment 1, that social media is most heavily used in the late evening.
Browser use was relatively stable from morning until late night, except for dips around 3pm and 10pm. I observed the same for email use, which declined gradually from late morning onward. The use of time & weather apps spiked around 8–9am, which makes sense since participants’ sleep diaries indicated nearly 60% of wake times were within an hour of 8:30am and over 75% of participants used the phone as an alarm (often a gateway to usage, they reported), similar to past findings that over 80% of people use a workday alarm [424].

Finally, entertainment apps were used more during the same morning period as well as in mid-afternoon and late night, while productivity apps showed usage peaks at points later in the morning, afternoon, and evening with a dip mid-day and dropping off past late evening. These patterns are similar to those found in prior work [54, 309], though shifted an hour or so later, likely because this experiment’s sample was younger and therefore trended later in terms of timing [428].
Weekly Usage Trends

I also found a distinction in the use of entertainment and productivity apps across days of the week. Figure 4.9 presents the percentage of usage for each category on each day, showing a reversed “scissor” pattern also found in other work [422].

For participants in Experiment 2, work days were Monday–Friday and free days corresponded to the weekend, Saturday–Sunday. At the beginning of the work week on Monday and Tuesday, I saw over 40% of productivity-based usage events occurring, while Friday and weekend days showed the least use of productivity apps — except for Wednesday, when only 8% of usage events were productivity-related. Inversely, Wednesday was the day when entertainment apps were used the most, followed by Friday and weekends. This mid-week dip resembles a common mid-week sentiment dip found in other work [4]; and in interviews, participants expressed experiencing a high degree of fatigue on

Figure 4.9: Use across the week of productivity and entertainment apps shown with standard error.
Wednesdays related to their class schedules — though further study is required to see if this mid-week effect is consistent inside and outside of other college populations.

### 4.3.2 Circadian Alertness Rhythms Reflected in App Use

Overall, these patterns replicated and expanded past findings and provided descriptive insight into types and temporal patterns of mobile application use. Yet, while this helps increase understanding of what individuals are doing with their phones over the course of hours and days, more work was necessary to understand why. I therefore next looked to chronobiology to add explanatory bite.

#### Usage Relative to Internal Time and Alertness

I first explored how usage patterns varied for different chronotypes. As mentioned, chronotype modulates nearly all biological functions [504], including alertness performance [109]. Simply put, earlier chronotypes are more alert earlier in the day, and later chronotypes function at their peak alertness later [212].

My comparison between the amount of usage events between early and late types across parts of the day suggested this distinction might be reflected in differing usage patterns, particularly of productivity and entertainment apps. Figure 4.10 shows these statistically significant differences in app use ($p < .05$ using Wilcoxon sign-rank tests) between early and late types, broken down by application category and time of day. Bars above (or below) the y axis indicate early types used that type of app at that time of day the indicated percentage more (or less) than late types.
Figure 4.10: Percentage increase (positive y value) or decrease (negative y value) in amount of usage by early types compared to usage by late types of productivity and entertainment apps across the day.

That is, in the earlier half of the day, early types used approximately 25% more productivity apps than late types and 19–29% fewer entertainment apps; while I observed the opposite effect for evening and night usage, when early types used 15–50% fewer productivity apps and 22–68% more entertainment apps than late types.

As mentioned, alertness exhibits well-known fluctuations over the course of a day [14]. The pattern of peaks and dips in alertness is roughly the same for everyone, but there are individual differences in the phase of these rhythms that are reflected by one’s chronotype. To align the phases of circadian rhythms for different chronotypes, temporal analyses of usage patterns can be shifted to a measure of time that is adjusted to take chronotype into account.

“External time” (ExT, also known as “clock time” or “local time”) is the number of hours that have elapsed since midnight (the midpoint of nighttime) [117]. “Internal time” (InT, also known as “body clock time” or “biological time”) re-
fflects the phase difference (i.e., the “phase of entrainment”) between time cues from the environment (e.g., the cycle of the sun) and the timing of an individual’s biological clock and is therefore computed as the number of elapsed hours since an individual’s sleep midpoint, $MS_{FS}$ (the midpoint of a person’s biological night) [426, 504]. That is, internal time is a corrected measure of time that reflects individual chronotype, calculated as:

$$InT = ExT - MS_{FS}$$

Reconsidering mobile app use in terms of internal time rather than external time, I now saw associations with participants’ innate biological rhythms of alertness (based on PVT performance — described in the “Daily Self-Reports” portion of Section 4.1.2, under “Alertness EMA”). Productivity and entertainment apps specifically showed the strongest associations with performance of all app types in Experiment 2’s dataset. Specifically, I found a strong positive correlation between performance and productivity app usage ($r = 0.52, p < .001$) — that is, higher alertness performance was associated with more usage of productivity apps. I also found an inverse relationship between performance and entertainment apps ($r = -0.31, p < .05$), indicating that lower alertness was related to increased entertainment app use, though the correlation was more moderate. Important to note is that I did not find any such strong nor statistically significant associations between alertness and app use when I did the same analysis using external time.

Figure 4.11 illustrates alertness together with usage of productivity and entertainment apps over the course of the biological day. Inspecting these trends beginning with the midpoint of biological night (hour 0 of internal time) showed how alertness levels gradually rose from the end of sleep through the wakeup
Figure 4.11: Temporal trends in application use (Usage) and alertness performance (Performance) across internal body clock time (InT). Usage axis is proportion (normalized to [0,1] scale) of all an app category’s usage events that occurred in a given hour. Performance axis is percent deviation from individual baseline of alertness measured in a given hour. Internal time axis is number of hours since biological midnight, and accompanying spectrum indicates periods of the biological day.

phase. During this same period, usage of entertainment apps was over 2.4 times higher compared to productivity apps. These findings resonate with the concept of sleep inertia, which can last for hours, reflects the transition period from sleep to full wakefulness, and is characterized by diminished alertness and vigilance in attention [428]. Nearly three-quarters of interviews supported this idea of an association between groggy wakeups and morning entertainment app use (e.g., “I’ll stay on the phone longer, browsing YouTube, etc, if I’m more tired.”)

Following this wakeup period, I found that alertness performance eventually peaked approximately 7 hours after sleep midpoint, which agrees with trends found in prior research [79, 183, 504]. At the same time, the use of productivity apps also ramped up and reached its own daily maximum, while the
use of entertainment apps fell to one of its minimum levels. Both the well-studied mid-day alertness dip (during which performance is known to drop [79, 341, 342]) and evening rebound [138, 287] were also observable in participants’ alertness patterns and aligned with a productivity app use dip and peak, respectively.

Finally, as biological night approaches, alertness is known to fade [504], and participants’ data showed this same trend in diminished alertness. (The outlier spikes at InT=1:00 and InT=23:00 resulted from sparse data since this period overlaps with sleep). In parallel, productivity app use also fell off while entertainment app use stayed more elevated. Accordingly, in interviews, participants commonly mentioned nightly habits related to watching videos or playing games (e.g., “Every time before I go to bed, I play a card game until I feel sleepy.”)

**Gauging Alertness Level from App Use Features**

I next explored how alertness may be reflected through additional usage features beyond the time of day an app is used. In Experiment 1, using technology for a longer amount of time was associated with procrastination, inattention, and lack of devoted concentration. In addition, switching among different tasks and computer windows has shown relations with capacity for sustained attention, distractibility, and boredom [309, 310]. Such findings suggest that the following metrics of app use duration, diversity, and switching may therefore be particularly relevant to alertness:

- Duration: Mean # of seconds per usage session during $T$
- Diversity: Total # of distinct apps used during $T$
- Switching: Total # of app switches during $T$
I calculated these features based on usage in a given hour window ($T$) surrounding an alertness measurement. To clarify, a usage “session” represents a period of interaction marked by unlocking the phone and is comprised of any number of app foreground events. I used Mann-Whitney-Wilcoxon tests to compare these features during low and high alertness states. Guided by prior research, I set thresholds for low and high alertness according to whether a PVT measurement was below or above that participant’s individual baseline — i.e., was a negative value in the range [-1, 0) or a positive value in the range (0, 1], respectively [504].

Participants’ overall durations of usage showed good agreement with durations found in prior work [54, 158, 541]. During periods of low alertness, I found duration of use was over 20% higher as seen in Table 4.6. Interviews agreed that usage became more “bottomless”, “stuck”, and “harder to get off”. I also found that participants also switched apps 33% more when alertness was low, though they did not necessarily switch among a larger set of distinct apps, as app diversity showed no significant difference between alertness states.

<table>
<thead>
<tr>
<th></th>
<th>Low Alertness</th>
<th>High Alertness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Duration</strong></td>
<td>103.4 seconds</td>
<td>85.8 seconds</td>
</tr>
<tr>
<td><strong>Diversity</strong></td>
<td>2.87 apps</td>
<td>2.82 apps</td>
</tr>
<tr>
<td><strong>Switching</strong></td>
<td>32 switches</td>
<td>24 switches</td>
</tr>
</tbody>
</table>

Table 4.6: Median values of usage features during low vs. high alertness. Significant differences in medians marked on variable name (*$p < .05$).
4.3.3 Connecting App Use and Alertness with Sleep

Lastly, I studied connections among app use, alertness, and sleep. Variation in performance is made most evident by sleep loss [184], with the largest effects on alertness, working memory, and cognitive throughput [288]. In addition to impairing performance directly, inadequate sleep is also associated with subsequent feelings of fatigue [269]. Conversely, extending sleep enhances learning and problem solving [509], with adequate sleep duration improving alertness, energy, and reaction time [232, 428].

Comparing sleep duration according to participants’ sleep diaries with their app use the following day, I found that less sleep was correlated with less productivity-oriented usage ($r = 0.43, p < .05$) and more use of entertainment apps ($r = -0.19, p < .05$), for both weekdays and weekends.

A more coarse grained measure of sleep duration adequacy also showed statistically significant differences. Specifically, considering (as in Experiment 1) sleep lasting 7–9 hours as “adequate” following established guidelines and prior research [88], I found participants used productivity apps an average of 61% more after nights of adequate as opposed to inadequate sleep (Cohen’s $d = 0.48, p < .05$), while entertainment apps were used 33% more on average after an inadequate amount of sleep (Cohen’s $d = 0.24, p < .05$).

Interviews provided qualitative detail about sleep loss and subsequent fatigue manifesting through increased usage of entertainment-based apps, which participants described as enabling “mindless”, “passive” interactions; while on the other hand, they associated feeling rested, energized, and alert with more “intentional”, “directed”, and “productive” usage.
Finally, recent studies continue to suggest links between nightly technology usage and sleep problems — particularly when it comes to usage delaying sleep onset and cutting into sleep time [68, 415, 485]. I found a mild relationship between the number of experienced sleep interruptions according to daily sleep diaries and the number of app usage events sensed between sleep onset and waking ($r_s = 0.46, p < 0.05$). This indicates app use data might help to assess sleep disruptions — though, the phone itself might be a culprit of disruption in the first place. In interviews, the majority of participants described turning their phone to silent overnight to avoid such sleep interruptions. However, even phantom notifications sometimes awoke them (e.g., “Sometimes I wake up as if I’m expecting something, like an email or a text, and will check my phone. I imagine that if I didn’t have technology, I’d have a sound sleep.”)

These results demonstrate how app usage can provide informative signals when assessing sleep duration, interruption, and associated feelings of alertness and fatigue. However, they also reveal a disruptive potential of mobile devices, which deserves careful consideration from researchers who leverage digital traces from these systems. I examine this tension further in the dissertation’s concluding chapter (see Section 6.2.1).
4.4 Discussion

Having presented a variety of results from both experiments, it would be worthwhile to now reflect on this research. In this section, I first provide a brief summary of these studies’ findings and their commonalities, then identify their limitations together with opportunities for future work, and conclude with broader implications for the domain-driven framework.

In Experiment 1, I focused on examining sleep, including sleep-related circadian disruptions and how inadequate sleep relates to changes in cognition and mood the following day. To begin, I analyzed temporal trends in the use of computer-mediated communication technology, along with associations between usage and sleep timing. Leveraging these usage patterns, I then inferred sleep onset and wake events with an accuracy comparable to approaches that are more obtrusive for users and less feasible to deploy on a mass scale. Applying this sleep sensing technique, I was next able to detect chronobiology phenomena such as the “scissors of sleep” and “social jet lag”. I further used social-sensor data to explore the impacts of sleep on variables defined to represent neurobehavioral functions known to exhibit strong circadian patterns and suffer substantially after sleep loss and interruption: attention, cognition, and mood. My analyses revealed significant differences in these variables following nights of adequate vs. inadequate sleep — specifically, that lack of quality sleep manifested in cyberloafing behaviors (based on an increased amount, frequency, and burstiness of technology usage the following day), diminished cognitive functioning (based on the expression of complex thought in text-based social media content), and more negative mood (based on sentiment analysis of that same content). Overall, my findings suggest that social-sensor data can serve as
a proxy to measure sleep timing, the extent of circadian disruptions like social jet lag, and the impacts of inadequate sleep on neurobehavioral functioning.

In Experiment 2, I explored daily functioning further. To scope, I focused on cognitive performance and even more specifically on alertness given how many other aspects of performance it underpins. Analyzing smartphone app use logs, I found that early and late chronotypes displayed inverse patterns of app use; that usage features such as duration and switching could distinguish periods of low and high alertness; and that app use reflected aspects of sleep including duration, interruptions, and subsequent fatigue. Notably, I demonstrated that adjusting analyses to biological “internal time”, as defined by the chronobiology literature, surfaced otherwise undetected correlations between alertness and app use, with individuals using productivity apps during their optimal performance periods and turning to entertainment apps when less alert.

Before considering the limitations of these experiments and the attendant directions for future work, it is worth relating their results to each other a bit more. In fact, the two experiments’ findings did align in interesting ways.

In both experiments, I found that less sleep correlated with particular usage patterns: elevated CMC usage in Experiment 1 and elevated entertainment app usage in Experiment 2, suggesting that CMC and entertainment software might have similar use cases and meet similar needs under some circumstances. Further, inadequate sleep was not only associated with a greater overall volume of CMC-based usage events the following day in Experiment 1, but those CMC usage events occurred at more frequent intervals too. Additional data captured during Experiment 2 further allowed me to find associations between inadequate sleep and lower alertness, which in turn related to longer usage sessions
that involved more frequent, defocused switching among apps. An interpretation supported by these data from usage logs as well as qualitative data from interviews, is that inadequate sleep is linked to subsequently diminished alertness and that CMC and entertainment apps are utilized as a means of cyberloafing to manage this experienced fatigue.

Because data about sleep, alertness, and CMC and entertainment apps were all collected in Experiment 2, it is possible to look at potential connections directly. Many of my reported findings for Experiment 2 focused on productivity and entertainment apps because they exhibited the strongest correlations with alertness, but social media and communication apps (i.e., CMC apps) together accounted for over half of that study’s usage events and sent some informative signals too.

To begin, I found usage of these CMC apps was elevated during biological morning, and nearly all participants in Experiment 2 described using social media apps as a way to “ease” themselves into the day (e.g., “To wake myself up, I’ll have to look at things on the phone like Facebook or Tumblr.”) Also notable is that these participants’ morning classes often fell within their phase of sleep inertia; and the majority of participants described using their phones in lectures when tired, bored, or unable to concentrate, for instance to help keep themselves awake — which they explained was particularly necessary for morning classes (e.g., “In morning classes, I have less attention and am very tired so I’ll browse the phone. Using tactics like social media, I focus on the screen to try to keep my eyes open.”) As presented earlier, participants in Experiment 1 described a similar “tactical” use of CMC to manage fatigue (e.g., “If I’m more tired, I’m less able to pay attention in class and more likely to use phone to avoid falling asleep or get bored
more easily.”) Given the negative impacts on learning associated with this type of distracting technology usage in the classroom [207], such findings suggest that the learning impairments associated with lectures being scheduled at biologically unsuitable times may be further compounded by this compensating phone usage.

Then throughout the day, participants in Experiment 2 appeared to continue turning to CMC apps when experiencing low alertness (e.g., “I go for apps that don’t require much mental energy when fatigued. Facebook, Yik Yak.”), including during the mid-day dip when these apps were used more than any other category. Reaching the day’s end, I found CMC apps also interplayed significantly with behaviors before and during sleep, with over 50% of sensed sleep interruptions corresponding to social media app use. When describing the amount of time between phone use and sleep in interviews, half of Experiment 2’s participants reported that usage was immediately before sleep and all but two reported it was within thirty minutes; and they noted this usage was often related to CMC (e.g., “I use my phone directly before bed — Messenger, email, Facebook. Any notification.”) Such findings resonate with those from Experiment 1 regarding ways in which CMC interacts with (including sometimes interrupts) sleep.

Altogether, these experiments provided evidence that biological rhythms exert a strong influence on patterns of sleep and alertness, and they demonstrated how these relationships and other phenomena well-known in a domain of interest may manifest through technology use. As a result, these soft sensor signals provide opportunities for more unobtrusive, scalable, and personalized assessment and intervention.
4.4.1 Limitations and Future Work

Broadening Samples

As explained, individual differences in circadian variables can vary dramatically. For reasons provided in Section 3.4.1, I focused on college students as a vulnerable population to study; but patterns are likely different for people of other age groups or who have different roles and work responsibilities. It is also possible that users of non-Android phones behave differently.

Thus extending this work to larger samples would be a natural future direction in order to see how well these experiments’ findings generalize across a wider sample of participants (e.g., of more diverse age groups, genders, and chronotypes as well as individuals living with affective illnesses such as bipolar disorder, who could benefit immensely from technologies designed to support circadian rhythm stabilization). Similarly, extending the study to a longer time frame and to additional geographical regions would allow measurement of circadian variations over the course of seasonal and yearly cycles and across multiple time zones and latitudes.

Further, I did not explicitly control for characteristics like class schedules, course load, or a number of other factors that might exert an influence on participants’ observed behaviors. Future work could extend models to include additional characteristics of participants and their contexts. Similarly, it would be desirable to examine the effects of light — including light emitted from devices, especially given my findings regarding the use of technology at night. Light plays a central role in setting the biological clock and the timing of sleep and is also known to impact alertness. While the MCTQ contains a question
about daylight exposure on average, measurement of daily sunlight as well as artificial light could be incorporated into analyses by including a question in participants’ self-reported sleep diaries or by using data from a phone’s light sensor. Such consideration of light could be particularly valuable to consider depending on a participant’s geographical location or the time of year, when differences in light-dark patterns may impact measures. Finally, though more burdensome for participants, comparison with actigraphy measures as well as exploring the use of body or environment based sensors could also contribute to a more holistic representation of rhythms.

**Broadening Data and Analyses**

Future work can also expand the types of data collected and the analyses performed. Additional qualitative data from diaries, EMA prompts with open-ended questions, or more interviews could help to further unpack and explain my quantitative observations regarding relationships among technology use, sleep, alertness, and other neurobehavioral functions. Such qualitative data might also enable the identification of additional edge cases in order to incorporate more informative features and make sensing more robust.

Next, future iterations of Experiment 1’s sleep sensing algorithm can incorporate additional socially-computable signals, for instance from emails, social media platforms beyond Facebook, or other data that interviews suggested would be informative. This may not only improve sleep inference accuracy from relevant behavioral data; but it would also allow examination of if, how, and why individuals exhibit different behaviors in different technology-mediated social contexts and whether such variations relate to circadian factors.
Similarly, in Experiment 2, I focused on smartphone usage because of a number of advantages outlined in Section 3.4.2; however, relying solely on behavioral signals from phones may ignore useful data from other sources. In particular, I likely missed some use of productivity tools designed for devices with larger screens and better input methods (e.g., laptop and desktop computers), which would be desirable to consider in future work. Further, the mobile-based alertness EMA was a source of potential confounds as it can impact attention and usage [486], suggesting the value of exploring more unobtrusive measurement. Such alternative sensing strategies could also capture aspects of alertness that may manifest through non-use of technology.

Likewise, it would be worthwhile to study phone usage behaviors beyond instances of app use (i.e., an app taking the foreground). Many of my analyses focused on this initiation of app use since it is a well accepted, easily comparable, and reliably captured metric, as described earlier. Theoretically, this does mean a 15 second interaction and a 15 minute interaction could potentially appear the same way, as single usage events; however, the former case would likely look different in practice because other app use events would tend to happen during the additional 14.75 minutes, especially considering the amount of app switching I observed. Still, I recognize the value in future work to look deeper into other metrics of usage such as duration, revisitations, or chains of app use that may offer additional insights.

In addition, although my application categorization was broadly useful, it was unable to account well for apps that can be used in ways that map to both low and high alertness. For instance, I observed elevated use of entertainment apps in both low and high alertness states. A potential explanation suggested
by interviews is that “lightweight” games that do not require much mental energy (e.g., “mindless puzzle games”) are used primarily when fatigued, bored, distracted, and for procrastination; whereas games played when feeling energetic and alert tend to require more focus and attention (e.g., “active strategy games”). Similarly, correlations between participants’ email use and alertness, together with interview data, suggested that checking email may be more productivity-oriented earlier in the day and more about “killing time” or socializing later on and especially before bed. Modeling the actual behaviors enacted in apps, though challenging, might therefore give a clearer picture of the relationships between technology use, alertness, and biology.

Still, as the first studies looking at social media and smartphone usage as soft measures for studying circadian rhythms of sleep and alertness, I have obtained a variety of useful findings relevant to researchers interested in chronobiology, mobile sensing, or personalized technology design.

### 4.4.2 Implications for Domain-Driven Health Assessment

My overall motivation in presenting these experiments was to demonstrate how domain knowledge can guide a process of data collection and analysis — specifically, by informing important health determinants to analyze, avoiding computational costs of modeling unneeded features, guiding analyses, aiding interpretation of outcomes, and generating implications for effective interventions.
To begin, this domain-driven approach enabled me to capture data and extract metrics that are believed to have a greater impact on health than indicators I might have measured if I had followed an approach disconnected from the domain of chronobiology. For example, extant sleep technologies, like those I reviewed earlier in Section 3.2.3, have fixated on improving sleep hygiene in terms of metrics like sleep duration. The previous chapter’s inquiry into chronobiology revealed other indicators (e.g., the experienced degree of social jet lag) that are equally (or arguably, more) important to assess when it comes to evaluating sleep and overall wellness. Therefore in Experiment 1, I explored how technology-mediated social interactions and communication patterns could be leveraged to provide a means of assessing socially-rooted disruptions to circadian rhythms — disruptions that the chronobiology literature helped me recognize as key determinants of overall health but that I would have failed to include in an analysis that was guided instead by layman intuition.

Domain knowledge also helped me avoid unnecessary computational costs. After finding usage of social technology was coupled with sleep behaviors, I next harnessed this social-sensor data as a means of assessing sleep events and quality. My algorithm was able to infer sleep events to a level of accuracy comparable to (sometimes better than) prior work’s domain-disconnected techniques — plus, this leaner, more targeted assessment approach was less computationally expensive and privacy intrusive than the more data-driven strategies.

In addition, domain knowledge provided implications for alternative ways to analyze data. For example, my analysis of temporal trends in smartphone usage data in Experiment 2 revealed patterns that correlated with personal alertness rhythms, suggesting such data could be leveraged to passively sense per-
sonal performance. However, such relationships became apparent only when analyzed through the lens of *internal* time, a measure of time that inquiry into chronobiology made available to me. As with Experiment 1’s findings about disruptions, I would have stopped short of discovering Experiment 2’s findings about a link between alertness and smartphone use if I had followed an approach disconnected from the domain of chronobiology. Further, the notion that analysis of usage trends can be adjusted to take the internal time of the body clock into consideration has broad implications for PHI work that assesses human behavior across time. For example, a number of studies have aggregated social media data across external clock time to study daily rhythms in mood [187]. The same analyses corrected to internal time (i.e., using a timescale based on the biological timings of those social media users) would be intriguing; and given the extent mood correlates with circadian rhythms [232], I suspect the results would be particularly striking in this case.

Next, both experiments demonstrate how domain knowledge can aid the interpretation of findings. While my results aligned with usage patterns found in related HCI research, domain knowledge about the biology behind sleep and wake behaviors provided explanatory power. For example, in Experiment 2, I brought a biological perspective on daily performance patterns to the interpretation of how and why individuals use their smartphones in particular ways. A domain-driven approach enabled me to go beyond prior works’ descriptions of diurnal variations in app use in order to offer biological factors behind those variations. Higher level constructs such as cognitive engagement and boredom have been used in prior HCI work studying digital activity [309]; however, I learned from domain knowledge that such constructs are underpinned by lower level processes like sustained vigilance in alertness [145]. I expect that through-
out the day, people fluctuate to an extent across their boundaries of alertness depending on their tasks, interactions, and other contextual factors — but that circadian rhythm factors present a consistent limit to one’s cognitive functioning and hence are most helpful in understanding individual performance (and in turn, could be practically useful in the planning of cognitively demanding activities).

Further, these experiments show how a domain-driven approach can facilitate direct contributions back to that domain. For example, my scalable soft sensing techniques for sleep and alertness assessment could benefit the research of chronobiologists who express a pressing need to capture in-situ data from large populations spread across diverse locations and time zones [423]. This ability to detect and infer behavioral traits of circadian biomarkers opens up new opportunities to answer fundamental scientific questions about sleep and circadian rhythms in real-world settings. Thus another strength of a domain-driven assessment approach is that the outputs of analysis are relevant not only to the PHI enterprise at hand, but they can have broader ramifications too by contributing novel methods or knowledge back to the informing discipline.

Finally, beyond contributing computational assessment techniques and empirical findings, both studies’ domain-driven approach also allowed me to derive a variety of design implications for novel systems that may considerably enhance monitoring and intervention. In the next chapter, “Domain-Aware Intervention Design”, I present in more detail the research I have undertaken to pursue those ideas.
CHAPTER 5
DOMAIN-AWARE INTERVENTION DESIGN

In the previous chapter, I showed how domain knowledge can support the collection and analysis of personal data in order to extract meaningful insights about health. For that information to be useful to individuals, it must be communicated in an understandable, actionable way. Therefore, this chapter moves along to the remaining component in the personal health informatics pipeline and, correspondingly, the final piece in this dissertation’s domain-driven framework: engaging in domain-aware design processes to deliver user-facing feedback. (See “Domain-Aware Intervention Design” portion of Figure 2.1).

This design stage begins in a phase of planning, where the goal is to gain an initial sense of what the PHI system will do and how. To better understand a system’s “requirements”, designers can gather information through several typical methods. For example, behavioral or social science theories can provide high level design principles applicable to a broad class of systems (e.g., using Fitts’ law of human psychomotor behavior to inform the design of pointing and input devices or using social science concepts about social capital or self-presentation when designing online communities) [441]. One might similarly look to sources of “textbook” design knowledge about best practices and broad rules of thumb (e.g., style guides or heuristics).

Also important at this stage is engaging with anticipated users, for instance through questioning (e.g., via surveys, interviews, or focus groups) or observational fieldwork in order to ensure a system meets their envisioned needs and respects their extant practices. Earlier, I described a recent shift in medicine toward a more person-centric model of care aimed at meeting individual health
needs and preferences. Similarly, a core principle in human-centered design is that technology should be user-centric, designed around users’ needs in order to create a positive user experience that demonstrates a deep understanding of their perspective.

A domain-aware approach to planning additionally integrates knowledge acquired during other stages of the framework (e.g., information gathered during domain inquiry or empirical insights generated during health assessment). In Section 2.3.2, I discussed the merits of informing design work with such domain knowledge in the context of PHI development, for instance, to ensure a system is both targeting the factors that are relevant to a health outcome of interest as well as accommodating the idiosyncratic characteristics that can impact how a person will respond to feedback.

Altogether, this information can then support a phase of ideation to generate implications for design, such as speculative guidelines for what feedback to present and how. A phase of building then involves the realization of these ideas, by creating low or high fidelity mockups and prototypes or even full-fledged systems where user models are instantiated. Finally, a participatory review process supports the evaluation of these built artifacts — information that can loop back and continue to fuel this ongoing design process aimed at regularly checking in with user needs and iteratively refining the PHI system.

Overall, these domain-grounded, user-centered steps support the design of systems that supply effective interventions as well as positive user experiences. In this chapter, I demonstrate such domain-driven design work in practice by presenting my research on creating chronobiology-aware PHI technology.
5.1 Chronobiology-Aware PHI

As described in Section 3.3, the chronobiology literature establishes that our circadian rhythms influence nearly all aspects of physiological and neurobehavioral functioning, including our sleep-wake patterns, cognitive performance, and mood. In addition, individual differences exist in these functions (e.g., the timing of sleep-promoting hormone secretions), as reflected by a person’s chronotype. Designing PHI technologies that are aware of such biologically-rooted, idiosyncratic characteristics opens up numerous opportunities for monitoring, stabilizing, and helping individuals work in better alignment with their innate biological rhythms — ultimately (ideally) to improve everyday life on a broad scale. The following subsections discuss particularly promising application areas related to sleep, daily performance, and emotional wellness.

The chronobiology-aware designs I present are of various fidelities — ranging from speculative design implications (e.g., unimplemented design ideas), to mid-level mockups (e.g., wireframes, storyboards), to working prototypes. In describing these designs, I indicate how they embody the dimensions outlined in Section 2.2.3 regarding format, delivery medium, attentional demand, room for interpretation, and level of personalization. Finally, while some designs (e.g., the calendar system) might not seem like conventional types of personal health technology, it is worth noting that the designs I present do meet PHI criteria — they use personal data to assess physical and neurobehavioral aspects of health and deliver feedback that can support individuals in self-managing wellness.

Regardless of the fidelity, embodied dimensions, or conventionalness of the designs, they illustrate domain-driven design practices, in that they reflect in-
tentional choices informed by available literature (e.g., chronobiology theory, clinically-validated interventions, condition pathology, etc.) as well as gathered evidence (e.g., the empirical findings presented in the previous chapter) and are grounded in user feedback that helps to verify a priori design ideas.

The following sections demonstrate various phases of the domain-aware design process. My descriptions of systems for supporting sleep (Section 5.2) focus on the planning and ideation phases, while my mid-fidelity mockups and prototypes for biologically-friendly productivity technology (Section 5.3) demonstrate more of the building phase. Finally, the participatory design and deployment of MoodRhythm (Section 5.4) showcase the building and review phases of a more high-fidelity, full-fledged system and also illustrate how insights gathered during this review can loop back into the iterative design cycle, in order to refine the system further and inform future design opportunities.

5.2 Chronobiology-Informed Sleep Support

As described in Section 3.2, given the current prevalence of sleep problems and comorbid conditions, PHI developers are keen to measure, assess, and improve various aspects of individuals’ sleep habits. However, (1) generic sleep hygiene systems that do not consider circadian factors are missing the full picture, and (2) interventions that only target sleep disturbances may merely be treating the symptoms of a misaligned biological clock rather than helping to address the roots of circadian disruption. In developing chronobiology-aware sleep technology, these two issues stand out as initial directions to pursue.
5.2.1 Personalizing Sleep Hygiene Recommendations

For patients who complain of sleep problems, clinicians often recommend that they should work to improve their “sleep hygiene.” Sleep hygiene is defined by the National Sleep Foundation as the “practices that are necessary to have normal, quality nighttime sleep and full daytime alertness.” Examples of these recommended practices include getting 7–8 hours of sleep per night, avoiding exercise within 3 hours of bedtime, and generally limiting caffeine especially within 4–6 hours of sleep.

Most of today’s sleep technologies are built with some subset of these sleep hygiene recommendations in mind. For example, BeWell computes a sleep quality score based on adherence to this ~7 hour ideal sleep duration [275]. “Got Sleep?” computes a similar score but factors in adherence to additional recommendations related to caffeine and alcohol intake, meals, exercise, napping, sleep environment, and electronics use [438]. As described earlier, ShutEye uses a smartphone’s wallpaper and lock screen to show a visualization of how likely it is that various activities including caffeine, exercise, and napping will negatively impact sleep at that point in the day [36].

These sleep hygiene recommendations are a good place to begin thinking about the design space of sleep technology at a high level, though they tend to take a relatively generic perspective. In this case, generic advice is likely better than no advice at all; but it would be desirable to make these recommendations more personalized, given the highly individualistic nature of sleep requirements. Specifically, chronobiology research suggests that how we sleep is determined by multiple factors and contingent, in large part, on biological attributes such as a person’s genetic makeup, age, and gender.
For instance, genetics play a role in determining the effects caffeine has on a person, including the time at which intake will affect sleep [542]. Using an approach similar to the experiments I presented in the previous chapter, a PHI system could attempt to translate personal data streams into behavioral biomarkers of caffeine-related traits. This information could then help a system move away from more blanket recommendations (e.g., ShutEye’s “End caffeine consumption 8–14 hours before bedtime” [36]), in order to tune this window to a user’s predicted genetic response to caffeine intake and its personal impact on sleep.

As another example, every person has a distinct chronotype, as described earlier, with many individuals (especially young adults) falling closer to the late end of the spectrum. Therefore, common recommendations for early sleep timings as well as maxims like “The early bird catches the worm” and “Early to bed and early to rise makes a man healthy, wealthy, and wise” perpetuate normative values that do not necessarily fit with everyone’s biology [426] — yet many systems supply sleep hygiene advice based on these generic perspectives (e.g., Jawbone UP’s “Early to Bed” goals encourage users to adhere to earlier and earlier sleep times, as mentioned earlier). Instead, chronobiology-aware systems could supply tailored timings for sleep onset and waking that better fit with an individual’s unique chronotype signature.

Similarly, recommendations could be further tweaked based on gender. Currently, sleep hygiene recommendations are not gender specific, even leading ShutEye’s designers to explicitly discount the need to consider gender differences: “Sleep hygiene recommendations are not gender specific, and thus we did not attempt to recruit an even number of males and females for the study” [36]. However, men and women are different when it comes to sleep; women’s
circadian clocks tend to run earlier (by about an hour) and shorter (by about 6 minutes) than men’s [141]. A PHI system providing advice about sleep onset and duration could therefore incorporate gender features into its algorithm, recommending timings that fall a bit earlier and last a bit shorter for female users than for their male counterparts.

Overall, such examples illustrate how sleep hygiene advice could be prescribed in more biologically-personalized ways. A natural starting point for chronobiology-aware sleep support is to design tools in the style of extant PHI sleep hygiene applications — i.e., suggesting the timing and duration of sleep or computing a sleep quality score. But instead of providing one-size-fits-all guidelines, sleep-related recommendations would be tuned based on a combination of any available information about biological traits (e.g., to provide a genetically-appropriate caffeine cut-off or to recommend sleep timing and duration advice that factors in chronotype, age, and gender). Responses from post-study interviews in Experiments 1 and 2 indicated that all but two participants would find this information meaningful.

Moving beyond timing and duration of sleep, metrics of sleep quality could also be made more chronobiology-relevant. This dissertation’s inquiry phase revealed other indicators that chronobiologists agree are equally important to assess when it comes to evaluating sleep and its impact on overall wellness, including sleep inertia, social jet lag, and sleep debt. Furthermore, such indicators are often symptoms of a misaligned biological clock, which chronobiology-aware interventions could help to stabilize.
5.2.2 Stabilizing Circadian Disruptions

As described in Section 3.3.3, disruption of our circadian timing system (i.e., circadian disruption or circadian misalignment) is associated with a wide range of negative health impacts. Aging as well as neurological and psychiatric disorders can lead to a breakdown in normal sleep-wake cycles; and even for healthy individuals, modern living’s biologically-unsuited work schedules, social constraints, and late-night technology use can all contribute to circadian disruption [528]. Therefore, another fertile area PHI sleep support can target is “fixing the broken clock” — i.e., stabilizing circadian disruptions through lifestyle interventions. One immediate design strategy is to create applications that can facilitate the broad-scale, real-world deployment of circadian interventions that have shown promise in small lab-setups or in animal studies.

Adjusting Sleep Schedules

Consider interventions to reduce social jet lag — the discrepancy between sleep patterns on workdays and free days, with the former typically characterized by chronic undersleep and the latter by compensating oversleep. In a recent study aimed at reducing this mismatch and promoting more stable sleep routines, factory shift schedules were adjusted to eliminate highly disruptive shift assignments. That is, early chronotypes were not scheduled for night shifts and late chronotypes were not scheduled for morning shifts. After five months, sleep duration and quality increased, as did wellbeing ratings; and participants’ social jet lag was reduced by one hour [503]. Technology could help deliver such chronotype-adjusted (CTA) sleep schedules to a wide user base. Given the informing research was conducted with factory shift workers, this intervention
may be particularly well-suited to individuals from that population; but others could benefit too. For example, doctors might use their recommended CTAs when appealing for a shift change or students could reference their CTAs when making decisions about whether to choose morning or night classes.

For a person unable to do this much of a schedule overhaul, a chronobiology-aware system might instead target the reduction of sleep debt. Sleep debt (also known as sleep deficit) refers to the accumulation of undersleep that people typically experience on workdays, and its misaligning effects are similar to crossing time zones [91]. Considered a central factor behind a number of adverse health outcomes, sleep debt is associated with increased daytime sleepiness, fatigue, mental exhaustion, confusion, mood disturbance, tension, and stress; and the effects of chronic sleep debt (lasting at least ten days) are similar to experiencing total sleep deprivation [133].

Recent chronobiology research suggests that helping individuals take on schedules that prevent the buildup of chronic sleep debt may be a promising strategy for reducing the detrimental effects of ongoing, unavoidable disruption (e.g., due to fixed work or class schedules that are ill-suited to one’s chronotype). Specifically, by adopting a rotating work schedule that gave at least 24 hours off after each night shift, study participants (shift workers) were able to immediately recover from sleep debt incurred from that night shift [160]. For a late chronotype college student who must take an early class every, say, Monday, Wednesday, and Friday, a chronobiology-aware sleep coach could help plan a schedule that ensures activities on Tuesdays, Thursdays, and Saturdays begin late enough to prevent the accumulation of sleep debt. Such a system might also provide a user with chronotype-tailored napping schedules, which the same
study found were effective at compensating for the ill-effects of shortened sleep after a night shift (e.g., napping up to three hours before a night shift for early chronotypes [160]). This finding about the benefits of naps also reveals how current sleep hygiene recommendations, which sometimes recommend the avoidance of any napping more than 8 hours past waking [36], could be made less brittle by supporting naps during later portions of one’s day, as long as they are timed in a biologically suitable way to stabilize rather than disrupt sleep.

Finally, another effective way to support healthy circadian rhythms is to maintain regularity in certain sleep-related “anchors”. In particular, keeping mid-sleep (the halfway point between sleep onset and waking) consistent each night can help stabilize circadian misalignment [528]. Even if a person is unable to go to sleep at the same time each night or wake at the same time each morning (e.g., due to work, school, family, or travel constraints), a chronobiology-aware system could recommend sleep and wake times that minimize circadian disruption by keeping mid-sleep anchored as much as possible. For example, again consider the case of the late type college student with morning classes three days a week (M, W, F). Say that this individual normally sleeps at 12am, that early morning classes require her to rise at 6am, and that she sleeps until 8am on non-class days. A bit of arithmetic shows that her mid-sleep point is 3am on Sunday, Tuesday, and Thursday nights and 4am on others. To eliminate these mid-sleep fluctuations, a chronobiology-aware sleep advice tool could suggest this user shift, for instance, her bedtime to 11pm on the nights when she does not have class the next day and her wake time to 7am on her non-class mornings. A quick recalculation confirms mid-sleep stability at 3am across the week. While this example is simplistic for illustrative purposes, an adaptive system could handle more realistic sleep patterns with more variable daily sleep and
wake times by determining on-the-fly an optimal sleep schedule to maximize mid-sleep stability based on the current time, the user’s historical sleep-wake data, and her known future schedule constraints.

**Delivering Zeitgebers**

Next, tools can supply interventions that help people temporally structure activities beyond sleep in ways that reduce circadian misalignments. As described in Section 3.3.1, our circadian rhythms are maintained by a process known as entrainment, whereby a group of nerve cells in the human brain use external information to keep our body clocks synchronized with changes in our environment (e.g., the Earth’s 24 hour light-dark cycle). The term “zeitgeber” (zeit: time, geber: giver) is used to refer to these external cues. Light is the most dominant cue; but a number of factors including temperature, food intake, and exercise work as zeitgebers too. The following zeitgeber-inspired design implications consider how chronobiology-aware technology could deliver these circadian cues in order to minimize disruptions and promote stability.

Using light to stabilize the circadian system has been a regular practice in order to re-establish healthy sleep-wake cycles for sleep phase syndromes [32], people with travel-induced jet lag, and shift workers [72]. Light is also commonly used to improve mood for conditions such as seasonal affective disorder [305] and depression [527], including for patients resistant to antidepressant medications [42]. Recently, light therapy has been extended to other conditions with symptoms of circadian disruption as well, including to improve sleep, motor skills, and cognitive abilities in patients with neurodegenerative disorders (e.g., Alzheimer’s and Parkinson’s) [525, 536] and to improve sleep and day-
time alertness in aging individuals [435]. Healthy, young individuals can experience positive benefits from light therapy too, including improvements in sleep quality, alertness, and fatigue [506]. Recent research has also demonstrated the efficacy of morning blue light therapy in reducing social jet lag as well as minimizing the associated losses in sleep quality and daily performance [178].

This encourages the design of chronobiology-aware light therapy tools, which could assist individuals in getting exposure to either outdoor light or artificial sources at the times that would help restabilize their circadian systems. For example, automated smartphone notifications or calendar tasks might remind a person to adhere to a personalized schedule of exposure. Advances in screen hardware on personal devices (e.g., smartphones, tablets) might even allow applications to deliver clinically-robust light therapy sessions without the use of specialized equipment. Novel light-emitting wearables (e.g., glasses, hats, visors, scarves, wristbands) could be another portable option [405]. Similarly, chronobiology-aware homes and offices might adapt the intensity and wavelength of lighting settings to cue exposure at opportune moments for realignment; however, while this might be straightforward for personal “smart lamps”, designing smart lighting for larger environments is a challenge, especially if individuals of different circadian profiles are together in the room.

However, improperly timed exposure to light can exacerbate circadian misalignment [448]. In particular, light at night is a known circadian disruptor. Given that many of today’s popular electronic devices for reading, communication, and entertainment have been identified as a main culprit of such disruption [84], building chronobiology-awareness of their users directly into these devices could help both to eliminate circadian disruptions in the environment
and to move people back toward stabilization. Software applications seem to be the best approach to controlling a device’s emitted light [151], and some tools do exist to automatically dim or adjust a screen’s white-blue light at appropriate times of day (e.g., iPhone’s Night Shift feature). The “f.lux” application is similarly designed to match the light spectrum generated by a screen with the natural light spectrum of the sun at any given time. This can help reduce sleep disruptions — assuming, however, that we all sleep and wake by the sun. A personalized, circadian-attuned version of the software could have default settings based on our available knowledge about the circadian effects of light but then automatically calibrate (or at least allow users to manually specify) “sun-rise” and “sun-set” times that match biological sunrise and sunset and individual sleep-wake cycles.

While not as potent a circadian cue as light, temperature is also a major regulator of sleep timing and duration in humans. In nature, the daily rhythm of environmental temperature is tightly coupled with the rhythm of sunrise and sunset. A recent study on the effects of temperature on sleep in pre-industrial societies found that sleep onset coincided with a nightly reduction in ambient temperature and that waking occurred just before ambient temperature started rising for the day [545]. However, this cycle of temperature change is largely absent in the modern sleep environments of most industrialized societies, with insulated buildings and artificial heating and cooling systems. The study’s authors suggest that recreating temperature conditions that the human body would experience in the natural environment of temperate climates might be greatly beneficial in stabilizing our own biological rhythms. Existing sleep hygiene tools often do encourage users to keep their bedroom at a “cooler temperature”, though their suggested range is a bit loose (60–75°F); instead, a
chronobiology-aware heating system could more closely simulate the nightly temperature fluctuations experienced in natural conditions.

Another feature of industrialized societies is that food is always available for most people, who can eat any time they choose. However, animal studies suggest that eating too much food, too frequently, or at inappropriate times (e.g., during the day rather than night for a nocturnal animal) can lead to circadian disruption, particularly with respect to metabolic imbalances [17]. In humans, regularity in daily routines, including stable meal timings, correlates with higher subjective sleep quality [343]. Eating meals at times that reinforce our biological clocks’ innate oscillations may therefore be an effective lifestyle choice for maintaining healthy circadian rhythms [448]. Further, considering the public interest today in using technology to manage diet and weight, systems that provide chronobiology-aware mealtime interventions might pair well with other PHI applications for healthy eating.

Finally, there are opportunities for chronobiology-aware exercise coaches to improve sleep and reduce circadian misalignment. Exercise can accelerate stabilization of circadian systems that have become desynchronized [336] as well as minimize the circadian disruptions of shift work [144]. In older men, mid-day fitness training has been shown to improve sleep-wake rhythms [497], while older adults with insomnia have experienced improved sleep quality after doing moderate aerobic exercise in the afternoon and early evening [412]. Chronobiology-aware technology could take into account personal characteristics like age and gender in order to supply information about the type, intensity, timing, and duration of exercise in which a user should engage in order to maximize the circadian stabilization benefits.
5.3 Biologically-Friendly Productivity Technology

Rather than an immutable trait, cognitive performance is coming to be seen as a critical component in the overall notion of wellness, with performance optimization emerging as a new frontier in health [475]. At the same time, pressures to boost work output in today’s increasingly technological and always-on cultures often endorse a mindset that it is possible to maximize — even “hack” [375] — human performance in a way that would allow someone to sustain high levels of lasting productivity.

However, such a perspective toward optimizing performance does not account for both inter- and intra-individual variability in biological characteristics [504] — i.e., the “internal timing” of the body clock. Beyond sleep, biological clocks also influence our cognitive performance levels, which naturally rise and fall throughout the day [79], as described in previous chapters. Alertness, attention, reaction time, response inhibition, short-term and working memory, and higher executive skills all follow rhythmic patterns [49].

Relatedly, technologies aimed at supporting productivity are typically designed on assumptions that our capabilities over the course of a day are steady or could be made steady. Calendars, for instance, typically treat hours and tasks as commodities instead of helping people schedule in accordance with their own historical patterns of performance. Notifications arrive at any time of day or night on the sender’s schedule, not the receiver’s; and though there has been much research around interruption management [25], it tends to focus on minimizing disruption rather than whether a person has the biologically-regulated cognitive capacity to respond to a particular kind of notice.
A greater awareness of our innate biological rhythms could positively change the way we design such technology. For instance, tools might dynamically adapt to the idiosyncratic needs of their users based on their current or predicted levels of performance, which could in turn support improved productivity on a broadly deployable scale. In the previous chapter, I demonstrated how circadian rhythms of alertness, a cornerstone of cognitive performance [445], could be measured using smartphone application logs. In this section, I discuss design opportunities for chronobiology-aware technologies that build on such assessments in order to support personal productivity in a biologically-friendly way. Particularly promising classes of technology include those that raise (inter)personal awareness of performance rhythms, scheduling tools, and performance-predictive systems. Below, I consider each in turn.

### 5.3.1 Self- and Social-Awareness

To begin, this line of chronobiology-aware designs aims to help people gain a better understanding of personal performance rhythms. Since people may not be aware of their alertness in the moment [139] nor have a good sense of why and when they experience alertness fluctuations, such systems could help individuals become cognizant about personal characteristics into which they would otherwise have little insight. Such self-knowledge could in turn empower a person to make more biologically-informed decisions when it comes to productively managing activities — or increase personal empathy and one’s capacity to understand, accept, and even embrace productivity dips.
Given that part of the goal here would be to help people learn about and come to appreciate their rhythms, a desirable strategy is to convey personalized alertness levels in an open-ended way that leaves room for interpretation and self-driven decisions. In the following designs, I therefore represent alertness levels through a peripheral background that uses a customizable color scale.

By default, a brighter, more yellow saturation corresponds to higher alertness while a faded blue-gray color corresponds to lower alertness. My initial design used a red color scale (i.e., more or less saturated shades of red). However, informal feedback-gathering sessions indicated that red was confusing or put people off; for example, individuals considered high alertness a desirable state, but that clashed with negative connotations (e.g., danger) that they associated with red. I therefore adopted the current color scheme based on prior design work that found people associate bright yellow with liveliness and energy, muted blue with the opposite (e.g., low energy, calm, relaxation), and both with agreeable perceptions [245]. Collecting some follow-up feedback, I veri-

Figure 5.1: To provide peripheral self-awareness, this live wallpaper’s color transitions in real-time in accordance with the user’s alertness levels at that moment. By default, brighter, more yellow saturation corresponds to higher alertness while a faded blue-gray color corresponds to lower alertness.
fied that the new scheme fit individuals’ mental models of how high versus low alertness would look as colors, and I confirmed that neither end of the spectrum was upsetting in the way the red scheme sometimes evoked negative emotions.

I am exploring how individuals react to this idea across several different media. The smartphone version, illustrated in Figure 5.1, instantiates the background in a live wallpaper whose color transitions over the course of the day to display the user’s real-time alertness information. I am also instantiating this feature in a chronobiology-aware calendar, as shown in Figure 5.2. Compared to the smartphone wallpaper’s temporary, moment-to-moment view of alertness, the calendar background presents a more holistic overview of entire days,

Figure 5.2: This chronobiology-aware calendar background scaffolds self-awareness of personal alertness levels, which are represented using a customizable color scale. Visual indicators on events provide an at-a-glance sense of whether scheduling aligns with personal alertness at that time.
weeks, or longer time periods and also enables a user to see archived views of past alertness patterns. In addition, visual indicators on calendar events reflect the required alertness of that event (using the same yellow–blue/grey color scheme) in order to provide an at-a-glance sense of whether the timing of an event aligns with personal alertness levels at that time.

This same information can also be communicated in ways that more intentionally expose it to other people, for instance through personal “beacons”: objects that are located in one’s environment (e.g., a desk ornament or room light) or used as a wearable charm (e.g., a necklace, pin, bracelet, ring, etc.), as illustrated in Figure 5.3. By providing peripheral cues about internal states, such displays could not only increase personal self-awareness but could also socially communicate normally invisible characteristics, for example to create a shared awareness of alertness profiles among co-workers.

Figure 5.3: The color of personal “beacons” transitions over the course of the day to display personal alertness levels in a way visible to other people. Beacons could be objects placed in the environment (e.g., a desk ornament, lamp, etc.) or wearables worn on the body (e.g., a necklace, pin, bracelet, ring, etc).
Such shared awareness could also be valuable in a family or other co-living setting (e.g., roommates). Viewable from a digital display in the home and synced to personal devices, a “family portrait” interface as illustrated in Figure 5.4 could show the individual alertness and sleep-wake patterns of each household member as a way to coordinate scheduling and foster empathy for each other (e.g., increasing a family’s tolerance of their teenager’s late sleep-wake schedule by improving their understanding that it is driven by biology, not laziness).

Figure 5.4: This “family portrait” (or “roommate portrait”, etc.) interface syncs to personal devices and uses a shared digital display in the home to deliver chronobiology information (e.g., performance and sleep-wake patterns) about household members as a way to coordinate scheduling and improve interpersonal awareness and empathy.
5.3.2 Scheduling and Activity Management

In this series of designs, I consider how technology for scheduling activities could take into account the cognitive demands of those activities and the circadian profiles of involved individuals. If the designs described in the previous subsection are in fact effective at improving people’s awareness of their personal rhythms and the substantial impact these rhythms can have on functioning, it will hopefully increase users’ receptivity to such scheduling suggestions and motivate the adoption of this more prescriptive form of feedback.

For example, the chronobiology-aware calendar introduced above could assist with scheduling cognitively-intensive versus rote tasks, based on a user’s chronotype, sleep-wake patterns, and historical alertness rhythms. In the prototype I am building, pull-based assistance is provided in two ways. First, a user is able to specify an event’s chronobiology-relevant information (e.g., required alertness in Figure 5.5, or required physical exertion, etc.) in order to receive recommended times to schedule that event based on her circadian profile. An

![Figure 5.5: Specifying performance-related parameters (e.g., alertness, as shown here) enables scheduling recommendations that align with personal rhythms.](image)
attempt to schedule an event at an ill-suited time prompts an overt alert with a warning and a suggestion for a more biologically-optimal time.

In addition, the calendar’s “Fix My Day” feature offers rescheduling suggestions to better align events with personal performance levels. In the case seen in Figure 5.6 (where the background gradient is toggled off), the rearrangement eliminates overlaps between the user’s morning sleep inertia phase (as a reminder, a period of diminished alertness and functional ability) with a meeting and a workout that the user has specified as cognitively and physically intensive, respectively. That is, the calendar suggests alternative times for these events when it predicts alertness and athletic performance will be higher.

Figure 5.6: This “Fix My Day” scheduling assistant analyzes the day’s events in order to provide suggestions for rearrangements that optimize alignment with personal performance patterns.
It is worth noting that the calendar medium could also be well-suited to delivering several of the sleep and circadian stabilization interventions previously overviewed in Section 5.2.2 (e.g., by scheduling recommended sleep times or zeitgebers such as when to get light exposure, eat, or exercise). This calendar could also be extended by building in social features, which could go beyond traditional calendars’ focus on mutual availability when recommending time slots for group-based activities (e.g., corporate meetings or student study sessions) to additionally consider if most participants are likely to be closer to peak performance. The user models underlying such a system could further facilitate team management by helping to pair collaborators or form groups whose members are better synchronized in terms of chronotype and performance patterns.

To move beyond the calendar medium and support more flexible timeframes and high-level activities, Figure 5.7 illustrates mockups of smartphone and smartwatch clockface widgets that deliver these types of chronobiology-aware activity recommendations. Both interfaces provide a glanceable view of person-

![Figure 5.7: Smartphone and smartwatch clock widgets show glanceable views of personalized activity recommendations.](image-url)
alized suggestions for the day. As these designs are intended to give users a bit more agency in their decision-making, the system leaves activity suggestions more open-ended (e.g., to “work” or to “exercise” during time windows with fuzzy boundaries). If more hands-on guidance is desired, a user might enable the tool to deliver notifications with more specific directives (e.g., “go for a 30 minute brisk walk at 12pm on Stewart Avenue“).

Figure 5.8 illustrates an application I am developing to explore the delivery of these same recommendations through a more playful experience that hits different design dimensions. In the context of activity management, systems

Figure 5.8: Storyboard of a playful chronobiology-aware activity recommender that provides in-the-moment feedback through an interactive, haptic experience.
tend to have primarily pragmatic, optimization-oriented goals. The purpose of this design is to explore a more ludic approach, for cases when the user is as interested in fun as in efficiency and when the idea of productivity might relate less to work output and more to experiencing a delightful use of one’s time. Brainstorming sessions with users produced design ideas related to crystal balls and fortune tellers, which led to this interactive smartphone app that ranks customizable candidate activities based on a user’s alertness profile. When the user desires activity suggestions and is in a playful frame of mind, haptics like shaking afford engagement in this fortune telling experience.

5.3.3 Performance-Predictive Systems

Another fertile area is the development of adaptive tools that can automatically alter system behavior based on the user’s current alertness levels or predictive systems that can do the same based on inferred, future alertness levels (or other personal attributes or indicators that are salient to and valued by a user). For instance, productivity tools that block access to potentially distracting websites or software might adjust their restricted usage times to match those when it senses an individual should protect a period of high alertness. Chronobiology-aware mobile notification could similarly delay the delivery of potentially distracting interruptions until an alertness lull was detected.

Systems capable of momentary alertness detection as well as future alertness prediction could also help individuals make more informed choices for themselves. For instance, when contemplating whether an all-nighter will either be productive in the long run or instead lead to diminishing returns, the sleep-
decision-support application illustrated in Figure 5.9 would provide a user with feedback about the “net gains” (or losses, i.e., negative consequences) of staying awake, in terms of the impact the system predicts potential sleep choices will have on performance the next day.

Finally, impaired alertness performance can be a serious issue when it comes to safety, increasing the risk of occupational injury, industrial disasters, and vehicular crashes. Performance-predictive technology might help prevent such accidents. For example, a driving-intervention application could apply predictive models about a user’s alertness in order to determine whether to deliver a recommendation to avoid the road until resting, if assessed accident risk was too high. A more controlling instantiation in a smart car might go even fur-
ther, locking the car’s doors to prevent driver entry or automatically pulling the
car over. Moving beyond the soft sensing approach based on social media and
smartphone app data that I took in the previous chapter’s experiments, a PHI
system’s assessment component could swap in data from alternative passive
sensor streams more suitable to the applied context (e.g., acceleration sensors,
steering patterns, or radio usage for the driving context) that could be used to
continuously monitor performance and deliver just-in-time interventions.

Altogether, the designs presented in this section aim to help individuals
work in better alignment with their natural performance rhythms — either
through adaptive suggestions or by simply helping people become more aware
of their personal rhythms in the first place. An accompanying goal is then to
help optimize performance in a personalized, biologically-friendly manner. In
the face of heavy work pressures, individuals today are increasingly turning to
stimulants to artificially heighten performance and extend working hours [18];
however, a domain-driven perspective suggests that achieving consistently el-
evated alertness is unrealistic and contradicts our biology. I therefore feel it is
important to emphasize that my design ideas should not be posed as helping
people work harder, longer hours. Rather, by incorporating an awareness of
biological rhythms into research on performance and technology, my intention
is to support a vision of systems that are designed to help people adopt more
biologically-suited working schedules and realistic, healthy productivity goals.
5.4 Mental Health Management

For this dissertation, my case study research has focused on bringing chronobiology to the development of PHI technology that supports sleep, performance, and emotional wellness. In demonstrating my domain-driven framework’s health assessment stage, sleep and performance played the major role. My work on emotional wellness (including mental health) comes more into the picture now, in illustrating the framework’s design stage. While my design work in the contexts of sleep and performance helped demonstrate planning and ideation as well as building at the level of guidelines, mockups, or barebones prototypes, my design work in the context of mental health speaks to other aspects of the framework’s domain-aware design process: building at a high-fidelity, deployment and review, and iterative trips through the design cycle.

Specifically, Section 5.4.1 presents the building of the high-fidelity system MoodRhythm, including descriptions of how design goals developed during planning were translated into system features and how reviews with users and clinicians helped to further hone these elements. The work presented in Section 5.4.2 then attends to the boundary between review and planning, illustrating how the design cycle can be re-entered through continued user consultations (both large and small scale) to inform further refinement of a system’s requirements. The end of that subsection then moves into ideation by providing design implications applicable to MoodRhythm as well as self-monitoring technologies for managing mental health more generally. Finally, Section 5.4.3 continues to focus on ideation but demonstrates how the design process is not a unidirectional nor serialized pipeline — in this case, by revisiting the review phase after ideating in order to corroborate design ideas before undertaking any building.
5.4.1 Participatory Design of MoodRhythm

In my research on mental health, I focus on bipolar disorder (BD), which is recognized as one of the ten most debilitating illnesses worldwide and affects approximately 60 million people \[358\]. BD is characterized by episodes of mania and depression that are separated by periods of normal mood. Manic symptoms include elevated or irritable mood, hyperactivity, impulsivity, and sleep loss, while depressive symptoms include inactivity, fatigue, and social withdrawal. BD is chronic, and there is no cure. People diagnosed with BD expect to manage their condition for the rest of their lives.

As described in Section 3.3.3, substantial evidence shows that circadian disruptions are associated with a number of mental illnesses including BD. Stabilization of an individual’s circadian rhythms can be an effective strategy to reduce condition symptoms — meaning chronobiology-aware designs can play a major role in supporting self-care and condition management in this context.

Specifically, the Social Zeitgeber hypothesis suggests that the disruption of certain behavioral, social, and sleep-wake events can disturb circadian rhythms and, as such, is a causal factor in triggering mood symptoms in vulnerable individuals \[146\]. Tracking and stabilizing these routines is therefore considered a particularly effective non-pharmacologic, chronobiology-based treatment for diminishing symptoms of BD. The standard practice for tracking lifestyle regularity and BD symptoms involves paper-based diaries, such as the Social Rhythm Metric (SRM) \[344\], which is a central element of Interpersonal and Social Rhythm Therapy (IPSRT), a clinically validated psychosocial treatment for BD \[170, 331\]. However, nonadherence to the paper-based SRM is common, especially when concentration is compromised during manic or depres-
sive episodes; plus, the paper format hinders the synthesis of data into easily-digestible summaries and feedback.

One promising alternative to paper-based diaries is using smartphone technology for behavioral tracking and intervention delivery, given its high ownership levels including among individuals with BD [41]. Indeed, mental health treatment protocols that involve such technology are becoming increasingly acceptable and advocated by organizations such as the Institute of Medicine and the National Council for Behavioral Health. A smartphone application also seems a well-suited medium because it not only permits completion of the self-report parts of the SRM on a device that is (near) constantly in the patient’s possession, but a range of BD-relevant parameters (particularly, sleep-wake behaviors, activity levels, and social interaction) map well to smartphone sensors that could automatically detect such information, in ways I describe later.

MoodRhythm [312] is a patient-facing, cross-platform smartphone app built on the Open mHealth Architecture and developed by a large team of collaborators as part of a participatory design process involving clinicians, professional psychological researchers, and most importantly, individuals with BD. As illustrated in Figure 5.10(a), MoodRhythm helps patients track the five main behaviors (getting out of bed, starting one’s day, first social contact, having dinner, going to bed) used in the standard version of the SRM. Users can also add custom activities that help anchor their behavioral rhythms, set and track daily routine-related targets, and record notes.

Reviews with patients and therapists indicated that most BD patients completed their paper-based SRM entries in batches at the end of the day. MoodRhythm’s goal was to support momentary use of the app to increase pa-
tient awareness of daily targets and to limit the impact of memory impairments, poor concentration, and variations across stages of the illness. To achieve this, several paths were implemented to streamline the recording process, including using overt notifications and making it possible for patients to quickly record SRM events directly from the smartphone notification panel. Another way to promote momentary self-assessment could be to allow MoodRhythm users to record information directly from the lock screen — a lightweight self-tracking strategy HCI researchers are currently exploring, as I mentioned in Section 2.2.1.

The app uses color indicators to provide glanceable feedback for the current and past days about how well an individual is hitting behavioral targets and maintaining routine regularity. It also provides weekly feedback using color as well as natural language summaries, as seen in Figure 5.10(b). Another piece of clinical knowledge that the design of MoodRhythm takes into consideration is that individuals with BD have a higher sensitivity to rewards. MoodRhythm therefore rewards adherence with a variety of badges, as seen in Figure 5.10(c).
This feature excited all patients, based on their participatory design feedback (e.g., “I like the rewards”; “Yes, I LOVE badges”; “I got the badges. It gives me personal satisfaction that I have completed something. And, that’s good. It keeps motivating.”)

During a four week clinical pilot of MoodRhythm conducted by collaborators at the Western Psychiatric Institute and Clinic (WPIC) with seven individuals with BD, the app received very high usability scores from participants, who particularly appreciated the convenience of recording activities with the smartphone medium as well as the way the app provided feedback in real time [2].

5.4.2 Iteratively Refining Design Guidelines

With colleagues, I have conducted both large scale surveys (N=552) [354] and small scale interviews (N=10) [313] to inform an ongoing design process aimed at regularly checking in with user needs (e.g., what BD-relevant indicators would be helpful to build into MoodRhythm to further extend the SRM, does MoodRhythm fit into existing self-monitoring practices, and how can its design choices address the currently experienced challenges of self-tracking while being sure to preserve the perceived benefits). Through these studies, a number of insights emerged related to what people are tracking, how, and why.

What Are People Tracking?

Participants reported recording a range of indicators, with mood, sleep, finances, exercise, and sociability being among the most common (see Figure 5.11). I also found that these self-tracking practices could vary and evolve over time, often in parallel with an individual’s phase of illness. Currently,
MoodRhythm does provide mood, sleep, and activity monitoring, though these findings indicate finances would be desirable to support next, as excessive spending during manic episodes is known to cause financial repercussions [517]. Further, nearly half of participants in the large survey study reported that they also track various other items that are relevant to their condition, including medication, side effects, and doctor appointments as well as personal triggers and manifestations of symptoms such as caffeine and alcohol intake, pain levels, appetite, libido, suicidal ideation, and self-harm. 20% of survey participants noted tracking items like chores, pet care, leisure time, and recipes that are seemingly less health-relevant but that they explained help structure daily behaviors, which can improve symptoms. Such findings reinforce our design decision to provide support for tracking custom activities and are also a reminder of the idiosyncratic nature of symptom triggers and manifestations — further motivating the need for technology-based solutions to move in personalized directions.
How Are People Tracking?

To track such variables, participants in both studies reported using paper-based formats (e.g., journals, calendars, sticky notes) as well as digital tools (e.g., apps, wearables, spreadsheets), along with mental notes, and (less commonly) feedback from other people (see Figure 5.12).

![Figure 5.12: Percentages (y axis) of various methods (key) used to track specified indicator (x axis).](image)

Sometimes, participants reported elaborate tracking setups as necessary to accommodate personal tracking habits in ways technologies do not currently support. Figure 5.13 shows an example from our smaller scale study — an Excel spreadsheet that includes a simple macro to display personalized messages of encouragement based on patterns in the person’s data. This custom tool gave this participant a way to record the things most important to her. It also felt more privacy-preserving to create her own self-tracking tool — another crucial factor to keep in mind when making design decisions, given the stigma some BD patients attach to their illness.
Figure 5.13: Example of one participant’s custom tracking setup that captures personally meaningful variables, assesses daily status, and delivers messages of encouragement. Diary entries have been blurred to protect the participant’s privacy.

Why Are People Tracking?

The majority of participants in both studies expressed that their self-monitoring practices are beneficial in managing their BD, for example by improving interactions with their clinicians, by offering opportunities for introspection or identity-building, and by helping them gain a greater sense of self-compassion and acceptance. Participants also explained how the ability to recognize patterns or pre-cursors to symptoms let them take a more direct role in their own treatment as well as learn personal coping strategies that worked for avoiding or recovering from mood shifts. Though their clinicians usually introduced them to self-tracking by way of the paper-based SRM, participants described advantages to technology-based self-monitoring tools they had adopted independently, including that tools make data capture less burdensome, promote adherence, and provide visual feedback that encourages accountability.
Still, our participants encountered shortcomings with existing technologies, including a lack of support in capturing BD-specific indicators (e.g., well-known manic or depressive prodromes, mixed moods, etc.) or capturing at a level of granularity sufficient for BD management (e.g., reporting multiple moods daily) as well as problems with the usability of interfaces that make them too cumbersome to use, especially when challenged with symptoms.

Based on these findings, I have derived a set of design implications that MoodRhythm’s ongoing development continues to pursue. I offer these as guidelines in Table 5.1 for how tools can be designed to support positive aspects of self-monitoring yet overcome extant challenges, in ways that meet patients’ expressed needs and are more condition-tailored than generic tools.

<table>
<thead>
<tr>
<th>Design Implications</th>
<th>Needs Addressed</th>
<th>Representative Quotations</th>
</tr>
</thead>
</table>
| Deploy software across platforms, devices, and operating systems | Pervasive accessibility | • “The easier it is to access the program the more likely I am to use it.”
• “I like typing on my work computer but use my iPad and iPhone at home.” |
| Deliver proactive notifications | Promotes adherence to self-monitoring and behavioral regularity | • “I also just found eMoods for my smartphone. I am just now starting to like it but I need to set an alarm to do it.”
• “I am so chaotic I find it difficult to keep track of anything without help and prompting.” |
| Synthesize data and highlight patterns | Increases self-awareness and reflection | • “It has helped me see general patterns and to recognize personal triggers. And the more aware I am of the symptoms, the more I can do proactively.” |
| Provide encouraging messages and rewards or (after non-compliance) flexibility and forgiveness | Provides experiences of mastery and cultivates self-efficacy and self-compassion | • “My first few episodes I felt intensely guilty about failure. It was this intense guilt that made me feel suicidal. Recognizing symptoms of depression has allowed me to be much more forgiving during episodes.” |
| Integrate with clinical care via doctor-view interfaces, digestible summary reports, and modifiable settings | Facilitates improved acceptance, transmission, and interpretation of information by treatment teams | • “I find that the reports succinctly provide my doctors with a more accurate picture over time than what I can recall at any given time. It also helps me to create a dialogue with my providers other than the fact that I don’t feel well (mentally). It has also helped my providers to see symptoms and patterns that I wouldn’t have thought to mention in short, 15 minute appointments.”
• “I tried to share optimism [tracking app] but my Dr was confused by the graphs.” |
| Provide BD-oriented functionality | Allows tracking of indicators significant to BD management | • “I have not been able to find an app that I really like enough to use. One problem with charting apps is they don’t allow you to chart more than one mood a day. If you have rapid cycling the app is useless.” |
| Allow customization | Supports idiosyncratic circumstances, preferences, and goals including how individuals’ conditions, management practices, and needs evolve over time | • “Very tedious. Would prefer to customize the computer program to track routine, socializing, etc.”
• “I used the app Optimism for about 4 months… but it only gave feedback/patterns on a few elements. I then switched to an elaborate excel spreadsheet that provided better feedback/patterns, but it ran off my laptop & wasn’t ‘handy’/convenient for tracking when I have time. Now my day is HIGHLY structured & my mood very stable. I now track in my head, have daily google calendar reminders, keep a running list to monitor elements, and have alarms that help with sleeping, eating, etc.” |
| Implement user-friendly features | Alleviates hurdles to tracking, including during mood episodes | • “I used to use a mood app I found on my phone but it was confusing so now I just use the notes section on my iPhone or an actual paper journal.”
• “I find many of the mood tracking apps overly complex and overly rigid.” |
| Passively monitor and intervene | Reduces user burdens and supports continuous capture of data | • “I used to use a calendar on my wall (for tracking), but I had a long mood episode of more than a month and quit tracking.”

Table 5.1: Guidelines for designing self-monitoring technologies to support mental health management.
5.4.3 Implications for Preemptive Interventions

A finding prominent among all the engagements with individuals with BD is their high receptivity to ideas of “intelligent” technology-based tracking systems that passively mine behaviors, automatically detect and predict affective changes, and report feedback about potential symptom onset and appropriate coping strategies. An electronic version of the SRM still faces many of the same limitations as the paper form — the burden of remembering to track is still upon the user; and for individuals with BD, the reliability of self-assessment (especially during mania) is questionable for any self-report based instrument. Many of the participants expressed a desire to have tools relieve such self-management difficulties by proactively providing reminders to self-track or passively monitoring behaviors without any explicit user input required. By crisscrossing from ideation to review and back again, this subsection unpacks this promising design direction.

As mentioned, smartphone sensing capabilities do appear well-suited to the automatic detection of many key parameters of bipolar disorder that are objectively observable and do not require patients to actively reflect on an internal state. Based on criteria regarding cognitive and behavioral manifestations of manic and depressive episodes as well as participants’ own descriptions about ways in which their technology-mediated activities vary with symptoms, indicators of symptom onset could be passively collected using commonly available smartphone sensors or usage logging in the following ways: excessive/reduced activity from accelerometer and geolocation data; increased/reduced sleep from light sensor data, app use, and social media patterns; and increased/reduced social activity from microphone, geolocation, and social media data.
To verify these potential measures and identify others, collaborators and I conducted another survey study with 87 individuals with BD [314] to further inform how a person’s amount, timing, and types of technology use may exhibit measurable differences during manic and depressive periods as compared to balanced periods. Table 5.2 provides a categorization of the specific manifestations of mood shifts that resulted from my qualitative analysis of responses and that could be used to guide the design of MoodRhythm’s (or other systems’) automated intervention strategies.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Manic Manifestations</th>
<th>Depressive Manifestations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer-mediated communication (e.g., email, phone, text messages, Facebook messages, tweets, blog posts)</td>
<td>Repetitively (re)reading and sending messages — to the point it may be construed as spam — or writing an excess of online content</td>
<td>Avoiding email or dodging phone calls and texts</td>
</tr>
<tr>
<td>Social networking (e.g., Facebook, Twitter)</td>
<td>Checking social news feeds repeatedly</td>
<td>Avoidance of social media</td>
</tr>
<tr>
<td>Web searching and browsing</td>
<td>Obsession with performing web searches of managing and rapidly switching among multiple browser windows and tabs</td>
<td>Non-use of “safe” use (e.g., reading websites)</td>
</tr>
<tr>
<td>Streaming media (e.g., Netflix, Hulu)</td>
<td>“Binge watching”</td>
<td>Non-use of “zombie-like” watching</td>
</tr>
<tr>
<td>E-commerce</td>
<td>Compulsively online shopping</td>
<td>Non-use</td>
</tr>
<tr>
<td>Gambling and gaming</td>
<td>Obsessively gambling or playing games on computers or phones for hours — especially social games, high-action games, or multiple games at once</td>
<td>Non-use or solitary, “calming” games (e.g., solitaire)</td>
</tr>
<tr>
<td>Digital calendars</td>
<td>Excessively booking activities</td>
<td>Diminished use</td>
</tr>
<tr>
<td>Typing and audio</td>
<td>Faster, more careless use (e.g., more typos or more garbled speech)</td>
<td>Slower use</td>
</tr>
<tr>
<td>Technology-mediated risky behavior</td>
<td>Increased visitation of dating or pornography sites, sending X-rated photos, using more inappropriate and aggressive language in written content, or more risk-oriented web searches (e.g., to find tattoo parlors or research exuberant vacations)</td>
<td>None reported</td>
</tr>
<tr>
<td>Use timing and frequency</td>
<td>Late night use: excessively checking phone notifications, or paranoid checking partners' emails, social media accounts, or cell phone logs</td>
<td>Diminished use overall</td>
</tr>
</tbody>
</table>

Table 5.2: Variations in technology use identified as characteristic of mania and depression.

Related research has had similar success inferring mood for individuals with BD using sensor data from their phones (e.g., the Monarca system [175]), which suggests the feasibility of using mobile systems for symptom detection and prediction and, in turn, preemptive care and targeted interventions.

However, MoodRhythm differs from such previous work in two key ways: (1) the foundations of MoodRhythm’s design (the Social Zeitgeber hypothesis, the SRM, and IPSRT therapy) provide a theoretical and clinical basis
for the data collected and feedback supplied, as opposed to previous efforts, which are not influenced by clinically-established characteristics of BD, plus (2) MoodRhythm’s participatory design process engaged individuals integrally throughout development — including in long-term, in-situ use of the tool — in order to ensure its likely adoption, ecological validity, and ability to support real-life needs. In these ways, MoodRhythm’s design work has been informed by the chronobiology literature on BD, clinical knowledge about validated treatment strategies, and participatory engagement with potential users.

Earlier sections of this chapter exhibited the same domain-driven design ethos in other contexts. Altogether, a primary intention of this chapter has been to put forth and demonstrate an evidence-based approach to design that foregrounds a deep understanding of the scientific underpinnings of a targeted aspect of health alongside a sensitive consideration of the lived experiences, extant practices, and expressed needs of users.
CHAPTER 6
GENERAL DISCUSSION AND CONCLUSION

In the final chapter of this dissertation, I first look inward, summarizing previous chapters as well as contributions. I also reflect on my research through several discussion points. In particular, I consider the other side of the coin — challenging this dissertation’s selected assumptions, arguments, and approaches by surfacing tradeoffs important for system designers to consider. Technology can be the solution to our health problems! But what if technology is the culprit of those problems? System-driven approaches relieve user burdens and enable insights invisible to the unaided human eye! But what if those manual, unaided practices have personal value? Personally tailored interventions are more effective! But when does a snug fit become constricting? Data-driven methods are the nemesis of domain-driven modeling! But could they become friends, in a combined effort that compensates for each other’s limitations? As part of this contemplation, I not only acknowledge this dissertation’s limitations, but I offer up concrete strategies for balancing these tensions in order to minimize the risks and maximize the benefits of future PHI systems.

Then shifting my gaze further outward, I discuss opportunities for future work to build on the contributions of this dissertation. I first describe how PHI systems can move beyond the single-user model that HCI has largely focused on to date, in order to accommodate more socially-oriented health management practices that extend beyond the individual. I also consider ways in which the domain-driven framework I have presented in this dissertation could be applied in areas beyond health, identifying well-suited domains and using my own research to illustrate a concrete example. Finally, I outline several other fertile areas to pursue going forward and leave the reader with concluding remarks.
6.1 Summary of Chapters and Contributions

In Chapter 1, I set the stage for this dissertation. Motivated by the need for novel solutions to help address a modern crisis of chronic disease, I explained why personal technology shows promise as an effective strategy for supporting self-driven health management. In particular, I identified its ability to enable the delivery of care in a broadly-accessible and cost-effective manner, in large part due to its increasing ubiquity, technological capability, and user receptivity.

In Chapter 2, I described these motivating ingredients in more detail. I first overviewed how medicine, disease, treatment, and technology have evolved and entwined over the last ~200 years, from the roots of modern medicine to today’s “age of behavior change”. In describing the expanding role of technology in supporting personal health, I provided a synopsis of the terminology, definitions, and relationships among the prominent work in this area (e.g., “eHealth”, which encompasses “mHealth”, which intersects with “behavior change technology”). Because my work comes from an HCI foundation, I particularly focused on health-related technologies grown in that field (e.g., “persuasive computing”, “personal informatics”, and “quantified self”).

Drawing inspiration from the characteristics of these tools, I then constructed a tractable definition of the sort of technology I aim to advance with this dissertation research: Personal Health Informatics (PHI), a class of tools that support personal management of healthy behavior via three key mechanisms — personal data collection, analysis of individual characteristics relevant to a particular health outcome, and feedback to help a person gain self-knowledge and potentially change or maintain behavior accordingly. I next mapped the design
space of these tools, reviewed how existing systems embody such characteristics, and provided guidance about how to make choices within that design space based on the goals of a system and the anticipated needs of its users.

Next, I motivated a central argument of this dissertation, that domain knowledge can help drive this PHI development process in order to fully capitalize on the potential of these technologies. Specifically, I gave a working definition of domain knowledge (e.g., theoretical constructs, empirical evidence, practitioner expertise, user perspectives, etc.) and explained the benefits of domain-informed PHI along with the problematic aspects of more domain-disconnected approaches. Finally, this chapter primed the remainder of the dissertation by providing a high-level outline of my framework for domain-driven PHI development, with components for how to go about (a) domain inquiry, (b) domain-driven data collection and analysis to assess health, and (c) domain-aware design processes to build user-facing tools for supplying health-related feedback. Each of the following chapters then unpacked those components in more detail.

Chapter 3 explained the process of domain inquiry, which involves selecting a problem area; assessing the merit, feasibility, and appropriateness of a PHI solution; and identifying salient domains from which knowledge is gathered to inform the subsequent stages of development. To demonstrate this in practice, I used a case study from my own research, describing the value in taking a chronobiology-driven approach to develop PHI technologies that support sleep, daily performance, and emotional wellness. To motivate that work and simultaneously provide a knowledge resource that can support others’ research in that space, I provided an overview of the field of chronobiology and how concepts like circadian rhythms, chronotype, and circadian disruption relate to

192
the aspects of health I was aiming to support. Based on this knowledge, I then devised a soft sensing health assessment approach, scoped to studying sleep-related circadian disruption as well as cognitive performance in young adults.

In **Chapter 4**, I undertook the analytic plan that the inquiry process had informed. Using my case study to demonstrate that domain knowledge can guide how to operationalize constructs, define variables, and conduct analyses, I presented the methodology for two experiments exploring how soft sensed data could provide informative signals about sleep and daily functioning. This chapter then contributed the empirical findings from these experiments, which illustrated how phenomena well-known in a domain of study can be detected and assessed by analyzing a person’s technology usage patterns.

In the first experiment, I focused on examining sleep: diagnosing sleep-related circadian disruptions, measuring how inadequate sleep relates to changes in cognition and mood the following day, and overall exploring how social-sensor data can reflect these trends as part of passive health monitoring. In the second experiment, I dug deeper into daily functioning: building on chronobiological foundations about cognitive performance rhythms in order to explore and interpret a number of relationships among alertness, chronotype, sleep, and smartphone app use — data I found could be similarly leveraged for passive monitoring. Together, these experiments substantiated how a domain-driven approach to assessment helps ensure the most significant health determinants, indicators, and outcomes are modeled; conserves computational costs by concentrating analyses on those targets; aids the interpretation of findings; and contributes novel methods and scientific insights back to the informing domain.
Chapter 5 built on the findings from these experiments, together with knowledge drawn from the chronobiology literature and interactions with users, in order to plan, ideate, build, and evaluate designs for technology capable of providing feedback and interventions that target domain-relevant behaviors and health indicators. Continuing to work in the context of my case study, this chapter contributed chronobiology-aware design artifacts of various fidelities produced by moving through the components of the framework’s domain-driven design cycle. Specifically, these designs included guidelines for personalized sleep support tools that stabilize circadian disruptions, mockups and prototypes for biologically-friendly productivity technology that can increase personal and interpersonal awareness of rhythms and assist in scheduling, and a deployed smartphone app that helps manage bipolar disorder by minimizing circadian misalignments that can fuel the condition’s symptoms.

Which brings us to the current chapter, Chapter 6, where I now reflect on the research presented in this dissertation along with opportunities for future work.

6.2 Discussion

This dissertation generally promotes a technological, automated, personalized, domain-driven agenda. However, as with nearly anything, there are promises as well as pitfalls associated with each of these strategies. In this section, I take time to ponder the limitations of my research and engage in a sort of tradeoff analysis. I hope this discussion serves as the starting point of a conversation I believe is important to have about how systems can be made more aware of and sensitive to such tradeoffs. Each point of discussion and the ideas I share are
guided by my experiences in developing the PHI development framework and in employing it to conduct the case study I have presented in this dissertation.

6.2.1 Technology as a Double-Edged Sword

One of the central claims of this dissertation is that technology has the potential to transform health care for the better. I have been a PHI promoter, generally making the implicit assumption that the more technology is used, the deeper it is embedded into daily life, and the greater individuals’ adherence to their PHI systems, the better. I have good reasons for thinking this way — the positive aspects of technology use have been well documented, and the efficacy of a PHI system does depend in many ways on heavy usage. In particular, rich data streams of technology-mediated behavior can act as a window into a person’s daily life, increasing our fundamental understanding of the factors impacting a given aspect of health. Further, an abundance of data can also improve the robustness of modeling techniques to assess those factors and related health outcomes. In addition, an individual has a better chance of seeing and acting upon feedback and interventions if they are delivered through a medium with which that person frequently engages, which is especially important for information that is context- or time-sensitive.

What I have not discussed (and what the HCI community in general tends to discuss less — some, but less [272]) are the occasions when technology is not the solution to the problem but rather, it is the problem. I can provide examples from two contexts in my case study: sleep and mental health.
Despite recommendations to limit technology use before bed, multiple studies find that nightly usage regularly occurs and can contribute to sleep problems in several ways [68, 415, 485]: by delaying the onset of sleep; by alerting the nervous system (e.g., from exposure to light or mentally-stimulating content), which makes it difficult to fall asleep or achieve restorative sleep; and by interrupting sleep (e.g., when notifications arrive during the night) [193, 194]. Such light exposure and time-shifted sleep can also disturb our body’s circadian rhythms. I observed similar issues regarding sleep disruption stemming from technology (over)use or dependence in my own research.

From the interviews conducted with 29 university students during the case study’s Experiments 1 and 2 (described in Section 4.1), I found that technology usage bookended sleep for nearly all those individuals. Activities such as checking email, texts, and social media; playing games; and watching videos typically capped off the day for over three-quarters of participants, and all but one person reported sleeping with their phones on or next to their beds. During these interviews, the majority of participants described that they set their phones to silent overnight since it would otherwise produce sleep interruptions; however, even phantom notifications would sometimes awaken them.

From my work with individuals with bipolar disorder (BD), I similarly found technology use to be a common disruptor to sleep, which is especially troublesome in this context given that sleep instability can trigger mood symptoms for those with BD [202]. In the N=87 survey about technology usage habits (described in Section 5.4.3), nearly three-quarters of respondents reported keeping their smartphones bedside at night, 67% of respondents reported losing sleep due to late-night use at least occasionally, and 45% reported always or of-
ten staying online longer than they intend or saying “just a few more minutes” when online, including at bedtime. In open-ended responses, participants described having difficulty falling asleep even once they put the technology down. One individual’s comment articulated well the interlocking issues of addiction and sleep disruption: “I often feel anxious about not being connected, and I spend too much time online, especially on Facebook. I used to use Twitter and LiveJournal every day, but I’ve managed to cut back on those drastically, and I don’t overshare nearly as much. But I will definitely forego sleep just so I can check my email and Facebook.”

From this same survey, I also found that technology can have an agitating effect or even provoke symptoms for individuals with BD. 41% of respondents reported that technology could trigger their depressive or manic episodes either by causing social distress or by exposing them to upsetting content. First, just as offline social anxiety and isolation are known precursors of depression, individuals described that feeling socially excluded online is similarly disruptive to their emotional balance. Social media in particular was noted as a common source of such feelings, often causing a sense of being ignored or left out, jealousy, or missed opportunity, any or all of which would contribute to feelings of loneliness and hopelessness and ultimately lead to or amplify their depression. Second, respondents explained that negative content such as inflammatory or offensive posts and disturbing news stories could produce anxiety and paranoia and in turn, trigger or reinforce their depressive symptoms. Respondents reported that their mania could also be triggered — particularly by exposure to pornography or posts with sexual innuendo, and frustrating experiences with technology such as software crashes were also noted as agitating. Most prominently, participants reported that spending too much time with technology is overstimulating and can fuel a spiral into mania. Further, technology can not
only act as a gateway to symptoms for these individuals, but it can also make the ramifications of episodes more severe. For example, one participant explained that during a manic shopping spree, there is a limit to how many items can be carried, stores can be visited, and hence overall purchases can be made in the real world — but online, there are no such “protections” and the financial consequences can be devastating. Finally, I suspect some of the emotional and behavioral patterns tracking tools monitor could themselves become triggers if reported insensitively or even reported at all. For instance, instead of leading to useful behavioral changes, knowing about a social “off week” might fuel rumination on social inferiority, depending on how it is presented to the user, her mood status, and her current ability to act on such feedback.

Thus as technology continues to permeate into people’s personal lives, it is imperative to further explore the potentially negative aspects of usage, especially heavy usage. Emerging research is beginning to express a concern over an unhealthy connection it has found many people are developing with their phones in particular. One thread of such work focuses on whether mobile phones pose a hazard to physical health, for instance due to ergonomic issues or radiation exposure or due to safety risks in the context of driving. Another area of research is more concerned with the psychological risks of excessive use, which studies from around the world have found can include anxiety, depression, guilt, stress, and worry, often due to the wearing strain of a perceived expectation to be constantly available [319, 484].

Herein lies the tension. On one hand, attachment to our personal technology may be precisely what gives power to PHI solutions for sleep support and mental health management, given the data frequent interactions generate, the reli-
able and continuous monitoring made possible by such granular data, and the often-attended digital medium for delivering personalized feedback in a timely manner. On the other hand, prior work and my own studies have discovered ways in which technology can actually be the culprit of the health problems it is aiming to correct, for instance by disrupting sleep or exacerbating mental health symptoms. One participant with BD nicely articulated this tension as follows: “Technology is a double-edged sword. Sometimes it’s a trigger, sometimes it’s a caution sign, sometimes it’s a tool, and sometimes it can save your life.”

The question is then: There are a variety of positive aspects to regularly using technology, but it has its dark sides; how can system designers optimize the benefits while protecting against the risks? I see an opportunity for “protective” features whose priority is to dampen the detrimental impacts of problematic technology use or perhaps preemptively discourage usage patterns that may harm health.

An example of protective sleep technology is software that automatically dims a device’s emitted light at night. As mentioned, several such tools have recently come on the market (e.g., Night Shift on the iPhone, Twilight on Android, and f.lux on computers and tablets). In the future, a smart home could protect sleep as the night winds down by similarly adapting home lighting conditions, along with additional environmental factors that help regulate sleep (e.g., temperature, similar to the chronobiology-aware home heating system described in Section 5.2.2). Meanwhile, a bedroom’s sleep-protective sound system could begin playing pink noise, which a recent study finds can induce stable sleep and improve sleep quality [549]. Delicate delivery of such environmental changes would be key, for instance through subtle, gradual cues that send the mind and body a strong signal to sleep while ensuring the home remains comfortable.
In the BD context, a protective system might adapt to manic behaviors by artificially diminishing the performance of a person’s device in order to slow down interactions (e.g., to buffer against a destructive spending spree by more slowly loading search results, opening tabs, etc). Design work would have to proceed extremely carefully though, given my aforementioned finding that frustrating interactions with technology can themselves fuel symptoms. In addition, considering the idiosyncratic nature of BD episodes’ behavioral manifestations, a protective system might offer a range of customizable interventions. Then, given the impairment effects that these mood episodes can have on decision-making, the user could choose from these options during a period of wellness and provide authorization that the intervention be deployed when symptoms appear.

Such ideas seem more sensitive and humane than coarse approaches like simply blocking usage past a certain time at night to guard sleep or shutting off one’s device when manic patterns are detected. Still, they highlight another tightrope to walk: developing protective systems that are both effective and ethical so that protection does not, in practice, begin to more closely resemble domination or manipulation. This challenge is likely to be further compounded by the fact that some designers may see no moral dilemma here, instead actively arguing that such control is acceptable or even desirable if the ends are beneficial enough to justify the means. (Personally, I hope the offspring of protective systems is not a generation of “For Your Own Good” tools).

Providing user-adjustable settings and manual overrides is one way to help preserve user control. Or, some protective technologies might adopt an “advance directive” approach [90] similar to the idea above, where an individual
in a healthy state provides consent regarding the level of control she is comfortable with a protective technology exerting during periods of infirmity. In other cases, the implication may simply be to not design [38]. That is, a protective design implication might be that the best course of action is actually to remove technology in part or in full from the health management equation.

6.2.2 Manual and Passive Modes

This issue of autonomy relates to another design tension I encountered in my research: balancing the tradeoffs between user-driven versus system-driven approaches. As explained in Section 2.2, each component of a PHI system (capture, analysis, intervention) can be designed to be more or less “participatory” [453], where a system that is more participatory puts manual control in the hands of the user, while a less participatory system gives more responsibility to the system. Each end of this spectrum comes with advantages and drawbacks.

As overviewed earlier, the main benefits of user-driven approaches relate to the idea that deliberate, effortful health management activities can afford opportunities for personal reflection, insight, and growth. Such experiences can cultivate mindfulness, self-knowledge, fulfillment, and empowerment. This can be especially true for users with special requirements, like those managing serious mental illness, for whom the nature of the condition can interfere with identity construction and self-tracking can help in building a stronger sense of self. I found this to be the case in my work on bipolar disorder (BD).

BD can have a considerable impact on individual psychosocial development and in particular on identity development. The initial onset of BD often oc-
curs during the teenage years [392] — a period of significant biological, cognitive, emotional, and social development; plus further identity confusion can stem from contradictions in experiences of the self across different BD mood states. As a result, individuals with BD often struggle with difficulties in self-acceptance due to an inability to integrate these different experiences and formulate a stable sense of identity [221]. In my research, I found that individuals with BD often integrated manual self-tracking into personal and therapeutic practices as a means of building such identity. In the N=552 survey (described in Section 5.4.2), respondents expressed that more manual forms of tracking and sensemaking made them feel that they were taking a more direct role in managing their own health, which in turn helped them develop a more internalized locus of control — a perception of agency that they described as instrumental to recovery. They appreciated that this made them more “active patients” who do not “simply fill prescriptions”. Individuals explained that the control they could exert during self-tracking and the intentionality of that process also kept them calmer since it made their lives feel more structured, manageable, and purposeful. In addition, my findings suggested that not only did manual self-tracking boost confidence and self-efficacy but that the opportunities for introspection it afforded were able to foster individuals’ self-compassion and acceptance, which in turn helped them be kinder to themselves and make more nurturing and understanding decisions about their mental health.

At the same time, these individuals also remarked on concerns about the potential fallibility of manual tracking, especially when they felt their ability to self-assess was compromised during mood episodes. They also discussed how recall could be difficult not only cognitively but also psychologically, especially if the recalled experiences were negative or traumatic, as this would actually
draw their attention to and accentuate that distress. Considering a move away from manual approaches, nearly all participants in my studies had high receptivity toward more “intelligent” tracking systems that passively mine behaviors in order to automatically detect and predict affective changes. They also felt such passively sensed data could be more “honest” and less susceptible to personal biases, and a number of participants believed sensor-based systems would be able to provide insights that they as humans would find difficult to discover on their own. Still, several people acknowledged that while today’s devices may be able to enhance human capabilities when it comes to noticing patterns and making predictions, there is still a limit to the precision they can achieve.

A number of individuals conveyed an awareness of this tension between agency and automaticity, which is as follows: *System-driven sensing can relieve some of the difficulties associated with user burden and adherence as well as improve tracking accuracy and pattern recognition, yet such passive approaches diminish intentionality and opportunities for self reflection and can also interfere with agency building and identity development.* Especially as PHI systems move toward more automated forms of health tracking, it is important to consider this give-and-take.

My work suggests there is an opportunity to design systems that strike an effective balance between manual and automated activities in a way that fosters self-reflection while still relieving user burdens and remaining reliable. A number of designers have suggested a hybrid approach, and several passive sensing systems do allow users to contribute self-reported information. Beyond this, I think there is room to develop systems capable of moving back and forth between manual and automated “modes”. For example, a PHI system might favor manual mode during early stages of use or post-diagnosis in order to build
a foundation of agency, self-efficacy, and intimate understanding of one’s condition. During certain periods, however, the more automated mode would be activated (e.g., during mood episodes in the case of a PHI tool for BD). Then progressively with time, a system could transition toward more frequent application of the automated mode, as a user becomes more familiar and adept with self-management and would prefer more of a passive monitor that keeps an eye on her health without requiring much hands-on input. Just as people’s health management practices, preferences, and needs change over time, I believe there is value in creating more agile tools that can similarly adapt and evolve.

6.2.3 Avoiding Over-Personalization of Tailored Experiences

As described earlier, recent decades have seen the health domain move toward a more individual-centric paradigm, with increased personalization of diagnosis, treatment, and cure. Technology’s recruitment into the battle against chronic disease has similarly enabled more direct and personally-tailored delivery of care; and in Section 2.2.3, I specified personalization as one of the key dimensions in the design space of personal health informatics.

I did note that non-personalized approaches could suffice in some cases. For example, a few studies on increasing physical activity [71] or reducing weight [229] have found that a generic intervention is better than no intervention. Or, the personalized strategy may simply not be cost-effective. For instance, before smartphones and wearables came on the scene, the idea of providing personalized sleep care on a broad scale was considered highly impractical given the specialized, expensive, and far less accessible equipment required for individually-
accurate diagnosis and intervention [423]. Understandably, public health organizations therefore favored the specification of generic sleep recommendations over leaving people with no support at all.

In general, however, I espoused the benefits of highly personalized approaches and reviewed the various cognitive, psychological, demographic, or contextual factors that can vary between and within individuals and sway whether or not an intervention is effective. To further champion the personalization strategy, I also warned about instances when non-personalized feedback could be outright harmful, such as when biologically-generic sleep advice might introduce circadian misalignment if it does not fit the recipient’s chronotype, when a badly tuned intervention might trigger a relapse of symptoms for an individual with bipolar disorder, or when dealing with vulnerable populations in general [136, 161, 337].

Still, it is worth contemplating the potential downsides to very tailored experiences. Many of the conversations about this concern are in the context of search results and recommendation systems, including most recently, personal news feeds on social media sites like Facebook. Here I focus on two problematic aspects of over-personalization that are salient in the context of PHI.

One issue is related to a user’s perceptions of systems that feel “too personalized”. This can make a user feel uncomfortable and reject a system as “creepy” [250], especially if the user does not understand how it is making such specific personal recommendations [493]. A second concern is that personalization might lock people into feedback loops that reinforce their baseline attributes, attitudes, and behaviors (or the system’s model of those user characteristics) [278]. Related to the idea of “filter bubbles” or “echo chambers”, these digital
comfort zones can constrain the range of information to which an individual is exposed. Feedback that may very well be useful and well-received might be deemphasized or filtered out entirely by ranking algorithms that judge it as too dissimilar to a user’s profile, perhaps because that person has not seen, ranked highly, or accepted (i.e., acted upon) that feedback in the past. This loop can become reinforcing and self-fulfilling: the user becomes more like the image the system has of her, as she continues to absorb the information she is provided and adopt the behaviors she is recommended.

This sort of over-personalization has also been described as a form of bias, with algorithms perpetuating either (a) the beliefs that their creators knowingly or unknowingly bake into them from the start (e.g., about what is “worth” measuring”) [360] or (b) the statistical patterns they learn over time — patterns that are potentially discriminatory in that they reinforce existing, systematic health disparities that often align with differential outcomes elsewhere in society related to age, gender, race, sexual orientation, and socioeconomic status [48].

In fact, “implicit bias” is a well-known problem in health care, where clinicians’ subconscious attitudes or stereotypes can affect their understanding and decisions in ways that perpetuate health care disparities. Consider obesity as one example, where researchers have identified a reluctance of doctors to look beyond an obese patient’s weight when making diagnoses or suggesting treatment options, even if the pathway to reported symptoms (e.g., shortness of breath, pain, etc.) is an underlying condition entirely unrelated to how much the person weighs [393]. In seeking out domain knowledge in the form of clinical expertise, a PHI developer who spoke to doctors with such an implicit bias might construct features in a way that specified higher importance scores
(i.e., feature weights, no pun intended) to those related to a user’s weight information. Or, if that system’s algorithms were trained using medical records of patients with such implicitly biased doctors, the models might automatically pick up these patterns and similarly promote weight-focused diagnoses and treatments. Either way, a well-intentioned domain-driven system could end up propagating such bias.

Altogether, the tension is then: While personalization is associated with numerous benefits, how can a system avoid over-personalization, where feedback locks users into potentially creepy, constrained, or even discriminatory experiences? One idea is to design PHI algorithms that foster risk-averse serendipity. In the context of information retrieval, a system’s measure of accuracy is usually based on similarity — that is, the degree to which a piece of content delivered to a user is similar to what that person has already enjoyed (e.g., seen and rated highly) in the past. However, this is what locks individuals into cocoons that can dramatically reduce their chances of seeing something completely different. The notion of serendipity (well-discussed in the recommender systems community [264]) refers to a system helping a user break free from these similarity clusters by supplying content that is diverse, novel, and unexpected.

For example, a less serendipitous PHI tool to support smoking cessation might deliver feedback related to a standard set of cessation strategies such as personalized encouragement messages or reminders about adhering to one’s nicotine replacement therapy. Instead, a system designed to promote more diverse experiences might also provide fitness-based interventions (that the less serendipitous tool likely would have demoted as less relevant than traditional interventions in this context). Such an approach might take further inspiration
from health interventions employing “cognitive dissonance” (which generally see positive outcomes [172]) — i.e., motivating behavior change by helping individuals come to recognize the difference between their current status and envisioned self-image. For instance, the serendipitous smoking cessation tool might choose to suggest fitness activities that would initially be extremely challenging for a heavy smoker, in order to accentuate dissonance and motivate change. More generally, by promoting exposure to fresh and unexpected feedback in such ways, a serendipitous PHI system could help to challenge and expand a user’s comfort zone in order to encourage positive progress.

Various algorithmic approaches could be employed to produce these sorts of serendipitous feedback deliveries (e.g., after clustering results, using an item-voting scheme that selects items from different similarity clusters proportionally to their relevance scores [480]). Moreover, an important aspect of a serendipitous PHI system is its “risk-averse” nature, which builds in a measure of protection in order to guard against the delivery of content that is so mismatched with a user’s profile that it could risk dangerous outcomes. Domain knowledge could help in tuning the thresholds for a system’s risk parameter (e.g., based on the vulnerability of a user, the severity of her condition, and the health hazards linked to that illness). For example, a PHI tool for managing BD would be extremely risk-averse considering the potential for misguided interventions to be life-threatening in this context. Further, given the fact that consistency in one’s routines is immensely helpful in improving BD symptoms, designing for excessive novelty, unexpectedness, and irregularity would likely be undesirable in this case. On the other hand, a fitness PHI tool might consider the perils of divergent feedback to be more minimal, tuning its parameter in a way to instead minimize the chances of user dissatisfaction or abandonment — although...
within limits. For instance, recommending that a novice, unfit user (e.g., the heavy smoker mentioned above) do extreme interval training for the sake of serendipity or cognitive dissonance could indeed be unsafe.

In addition, PHI interfaces can be designed so that not just the content but also the presentation of feedback facilitates serendipity. For example, search engines sometimes avoid ranked lists in favor of grid views or organize results into category views that promote exploratory information discovery. Design choices in terms of how feedback is presented as well as the user’s control over that presentation can also help mitigate problematic aspects of personalization related both to bias as well as a user’s discomfort with feeling monitored. First, providing transparency or context about why a particular piece of feedback is being delivered can increase user satisfaction, as it helps a person understand how exactly her behaviors came to influence the feedback she is receiving [458]. Given that there are various ways that personalization can fail, such transparency can help a system fail gracefully. Transparency can also diminish a sense of creepiness, in large part because transparency builds trust — something particularly important in a health context, where users might have special concerns about confidentiality, privacy, sensitivity, and stigma. By exposing how the system works, transparency can additionally provide users with a sense of control.

In fact, the ability to control personalization mechanisms can be crucial to their acceptance [493]. A PHI system can give users control in various ways, for example by providing a means of deferring behavioral recommendations or through options to turn personalization on and off. The ability to turn off personalization can help assuage privacy concerns as well as combat bias; a variation might be to provide a setting that lets a user see feedback personal-
ized for a different profile (e.g., “what would I see if the system thought I was a different gender?”). Further, design choices like the name of a system or how its functionality is framed can change users’ receptivity to the information it provides; for instance, a user might be more accepting of a PHI fitness “coach” that supplies novel and unexpected (and sometimes imperfect) feedback than the identical system branded as a PHI fitness “recommender system” [40].

6.2.4 A Dual (Not Duel) Domain & Data Driven Approach

Another core tenet of this dissertation is the benefit of taking a domain-driven strategy when developing PHI technology. Indeed, I spent a considerable amount of effort describing the value of domain-driven PHI (e.g., see Section 2.3.2) and the trouble with domain-disconnected efforts (e.g., see Section 2.3.4). I particularly took issue with more data-driven, “black box” approaches where models are guided by patterns noticeable in the data rather than on theory, empirical evidence, or other validated knowledge, arguing that such models could be inaccurate, irrelevant, misleading, or even disrespectful of human complexity. In this Discussion’s previous subsection, I further pointed out that purely data-driven models might reinforce pre-existing biases in potentially unfair, discriminatory, and harmful ways. In contrast, I argued that domain knowledge could thoughtfully inform all aspects of system design from data collection, to analysis, to intervention, to evaluation. By guiding the important health determinants to target, avoiding the computational costs of modeling unneeded features, aiding the interpretation of outcomes, and enabling the refinement of existing theories, I do believe a domain-driven approach can drive the development of systems highly effective at improving people’s health.
That being said, domain-driven approaches can have limitations. First, domain knowledge may be biased, incomplete, or simply unavailable. Especially when working in an understudied or novel area (e.g., relatively uncommon health conditions), pre-existing scientific literature is likely to be sparse. In such cases, one can try to generate domain knowledge, for instance by engaging with the anticipated users of a system to determine practices or needs; however, lack of access to such individuals may be similarly challenged, especially if the population is small, remote, or marginalized.

In other cases, knowledge might exist but its relevance is questionable. For example, empirical evidence might have arose from a context very different from that of the current PHI project (e.g., many years ago, with a very different population, or under otherwise substantially dissimilar circumstances). In an increasingly diverse world of evolving health conditions, evidence can lag behind; and in today’s rapidly changing technological climate, cutting-edge research about user characteristics or usage tendencies may become obsolete even after a short period of time.

Furthermore, there will always be a gap between the health needs and idiosyncratic symptom manifestations of any one particular individual with aggregate findings (e.g., from a large-scale randomized controlled trial, or RCT) — if any such RCT has even been conducted. Just as I mentioned that some personalization strategies are cost-ineffective, the time and resources required to conduct RCTs for some conditions preclude them from ever being realistically conducted. For example, it has been estimated that it would require 127 RCTs involving 63,500 patients over 286 years to produce the evidence necessary to inform clinical decisions about Alzheimer’s [442].
In addition, there is a well-known publication bias against reporting null results, regardless of the quality of study design [143], which not only means that relevant knowledge may exist but be unavailable, but there could be a bias inherent to relying on the information that is available. Further, the flip side may be possible too: flawed results (e.g., from poorly designed or executed studies) do get published — unreliable knowledge that future researchers would mistakenly deem as a trustworthy foundation to build upon. While a well-seasoned researcher may be able to identify such red herrings, this may not always be the case, which relates to my next and final point.

Lastly, domain knowledge can be applied inappropriately. This dissertation’s framework, as I mentioned when I introduced it, is intended to serve as a way of thinking, a methodological process guide rather than a binding set of cookbook recipes. A domain-driven researcher should use the best available evidence to make informed decisions — but this process will still involve personal judgment to some extent (e.g., in order to determine what that evidence might be, whether it is in fact applicable, and if so, how to apply it). However, this takes time, effort, and practice. If a research team does not have the necessary skills, interest, or patience to rigorously scour, assess, and apply domain knowledge, then a shallow application can be equally or more problematic than a process absent of domain knowledge entirely. Researchers taking a domain-driven strategy but expending minimal effort are prone to confirmation bias — noticing, incorporating, and interpreting only the conveniently-acquirable evidence that confirms their pre-existing beliefs, previous experiences, or personal judgments, rather than alternative and potentially more relevant information.
For example, a common pitfall of HCI researchers applying the transteoretical model to develop behavior change systems is to consider only the model’s stages of change but not its constructs related to decisional balance or self-efficacy. Not integrating these aspects of the theory diminishes the potency of the full conceptual framework for designing a system, plus it makes it more difficult to evaluate outcomes of the system, for instance to explain why the system is or is not effective for some users (e.g., those individuals with or without strong self-efficacy) [206]. Or, as I mentioned in Chapter 2, reviews find that sometimes systems mention drawing on theory but supply few details or rationalizations [259, 382]. As I pointed out at the end of Section 2.3.3, this leaves it unclear in many cases whether those theories were actually the most suitable — or simply the most familiar and in vogue, given a cascade of prior work that had already been drawing on them.

Therefore to resolve this tension, we must consider: Are there worlds in which domain-driven and data-driven methods can peacefully co-exist — or even further, lead to PHI solutions more favorable than either could achieve independently?

I believe combining these two approaches, domain-driven and data-driven, provides a promising research direction: guiding practical, data-driven methods with responsibly-gathered knowledge from both the health and HCI domains (e.g., knowledge about the health condition being managed; needs, values, preferences, and practices of patients/users; advantages and limitations of available treatment options and relevant PHI tools; and applicable computational and interaction principles).
This process would be one of iterative exploration and experimentation, punctuated by critical reflection about what has been discovered, worked, failed, and how to proceed. Speaking back to the limiting case where domain knowledge is lacking (e.g., unavailable, incomplete, or biased), data-driven approaches can help to address such challenges.

In particular, data-driven identification of patterns can be useful in constructing hypotheses, which can then be evaluated and, over time, solidified into verified theories themselves. In addition, such patterns might have been imperceptible via human-driven appraisals alone. This means data-driven approaches can help researchers generate hypotheses and, eventually, domain knowledge that we would have otherwise found difficult to construct. Further, going back to the bias issue, these previously unrecognized patterns might correspond to societal inequalities currently in our “blind spots” that data-driven computation helps make noticeable.

Altogether, I look forward to a future where data and domain knowledge can live side-by-side in a symbiotic relationship that helps to advance PHI solutions and our current health challenges.
6.3 Opportunities for Future Work

My overarching and long-term research agenda is directed at supporting well-being through positive interactions with technology. Throughout the course of developing a framework for domain-driven development and conducting my case study research, I have identified several additional areas I believe it would be worthwhile to continue exploring. In this section, I discuss these opportunities for future work that I see as particularly compelling.

6.3.1 Moving From Personal to Collective Informatics

In this dissertation, I reviewed a number of tools developed in recent years to support health management, and I presented domain-driven guidelines I argued could positively impact the way we create future generations of such technologies. I particularly emphasized a focus on the individual — relaying visions of people at the center of their own care, using technologies that support self-driven health management via personally-tailored experiences. I purposefully included the word “personal” in my conception of “personal health informatics” to stress this focus. Most extant systems and implications share the same model and have been developed and deployed from a single-user perspective.

However, health management practices are frequently embedded in social contexts. From data capture to sensemaking, individuals enact a range of social practices. For example, in my work with individuals with bipolar disorder, I found that social feedback was sometimes used as a form of condition monitoring, whereby individuals would periodically report their mood to loved ones or
would depend on trusted connections to act as a sort of “human sensor” who would pay attention for warning signs of a mood episode [354].

Engagement with collected data (whether captured solo or socially) can extend beyond the individual as well. For instance, individuals can use personal records to build understanding among family and friends, share data with groups of peers (e.g., individuals with the same condition) for social support, or transmit data to caregivers to facilitate treatment oversight. I found examples of this in research I have done on PHI for pain management, where interviews I conducted with individuals experiencing chronic pain revealed that they would sometimes use their pain logs to gain empathy from their family members (e.g., by using personal data to “prove” they were experiencing severe pain) or to increase credibility when discussing treatments with doctors (e.g., by using data as “hard evidence” to substantiate intuitions about their pain’s fluctuations or their medication’s efficacy).

Likewise, in the N=552 survey with individuals with bipolar disorder (described in Section 5.4.2), about two-thirds of respondents reported using self-tracked data with doctors, psychiatrists, or therapists as a way to open and maintain lines of communication. Specifically, individuals explained that this personal data helped them relay symptomatic patterns in a more aggregate and accurate manner than they could have done without the assistance of technology. They also mentioned that data enabled them to more accurately recount behaviors and events, especially if significant time had passed since a prior appointment. Additionally, similar to the pain patients I met, these individuals with bipolar disorder felt that data gave them defensible evidence for discussions about treatment efficacy or medication adjustments.
I believe there is a considerable opportunity to design computing infrastructures and interfaces that can better support such collective practices across all phases of health management. Some technologies do include social features, but they are typically geared toward comparing performance with other users (e.g., through leaderboards or competitions like Fitbit’s “Who can get the most steps this weekend?” challenge) or providing group discussion spaces (e.g., online health fora to seek information or social support). Peripheral public displays support interpersonal awareness to some extent (e.g., the designs I presented in Section 5.3.1 to promote social reflection about an individual’s personal circadian rhythms), but such work is still nascent in the context of health management. Going forward, novel collective informatics technologies could significantly enhance even deeper social engagement with the collection and sensemaking of others’ health data, in the ways my own research experiences have demonstrated individuals and members of their support networks desire.

As with the design work I presented in Chapter 5, user-centered participatory design processes could be undertaken to (1) more fully understand collective informatics practices, including who are the involved parties (e.g., care providers, family members, close friends, extended networks), how they engage with each other, and how technologies can facilitate these interactions and to (2) develop innovative collaborative computing architectures, data representations, interaction principles, and interface designs that meet these social needs. Such developments could help in scaffolding social support networks that play important roles in health management, increase the accuracy and utility of tracked personal data (e.g., by enabling a trusted circle of stakeholders to participate in its collection, verification, and analysis), and provide caregivers with new types of information that might support more effective treatment.
There are also a number of critical issues in this space that must be identified, documented, and designed for. For example, applications for collective forms of data capture, sharing, and sensemaking must be sensitive to data management (e.g., data ownership, privacy, and access) as well as how to protect the autonomy of an individual when a group of people come together through a mutual interest in his or her wellness. Overall, such responsibly designed sociotechnical systems could lead to a number of positive impacts at individual, group, and societal levels.

6.3.2 Applying the Framework in Areas Beyond Health

While this dissertation has demonstrated the utility of a domain-driven approach in developing PHI — personal health informatics — an important question remains regarding generalizability: Can a domain-driven technology development framework be applied in domains beyond those related to health?

As described in Section 2.2, PHI is essentially a subclass of personal informatics tools — the “H” is added precisely to distinguish PHI from the broader definition of personal informatics that can encompass a wider range of personal data. Seminal personal informatics literature (e.g., the publication that coined the term) has identified a number of domains where such tools have been or could be applied. Many are related to some aspect of physical, cognitive, or emotional health, as I have explored in this dissertation. Others that stand out include sustainability, information seeking, and online contribution. Examining the generalizability of the domain-driven framework in the context of these domains therefore seems like a reasonable thing to do.
To begin, academic and industrial HCI researchers have developed various tools to promote sustainable behaviors (e.g., for energy consumption, water usage, transportation, and waste disposal). One line of work studies the design of “eco-feedback” technology, which uses personal devices (e.g., smartphone apps) or home-integrations (e.g., sensors and lights on faucets) to measure and increase environmentally-friendly behaviors. For example, LCD displays like the Energy Orb have been used to display daily energy usage levels in a descriptive, peripheral way, while other systems like Kill-A-Watt or Energy Detective have more prescriptively provided persuasive prompts to reduce usage during peak hours when energy costs are high [205]. Similarly, the Shower Calendar is a prototype system that displays personalized water consumption information via a screen or projection located in the bathroom in order to foster awareness, family competition, and ultimately reduction of water use [276]. Systems like the Electricity Portal provide a more societal-level (in this case, city-wide) portrait of energy consumption that uses household-level feedback, incentives, and social comparisons to promote conservation [150].

I would argue that such technologies are very well-suited for a domain-driven development strategy, as seen by walking through the components of the framework. First, it is possible to identify a rich body of knowledge that provides insights into why people do or do not tend to engage in environmentally responsible behavior (e.g., from individual characteristics like altruism or fiscal concerns to group level considerations like societal norms) from domains including education, economics, philosophy, psychology, and sociology [173, 174]. Further, while PHI systems target health behaviors and systems like Energy Detective target conservation behaviors, both are rooted in behavior change [69]. This makes many of the domain-driven development stages quite
similar: drawing from behavioral theories (e.g., goal setting, social comparison, and positive/negative reinforcement have been employed in this context of sustainability technology [472]) in order to inform how to collect, model, and provide feedback about those behaviors, ultimately in order to increase opportunities for reflection and positive change.

The next application area, information seeking, is less analogous to behavior change yet a domain-driven approach still seems applicable. While most information retrieval systems use an inductive approach (i.e., devising user models by analyzing patterns in users’ interaction data) [292], a few have shown it is possible to take a more deductive, domain-driven approach. Specifically, such work has utilized established behavioral theories (e.g., information foraging theory or economic utility theory) in order to guide their modeling of users’ search strategies and information goals and inform the development of more personalized search systems (e.g., for images [292] or work-related information like emails or documents [67]).

As a final example, I have had success at applying the domain-driven framework in the context of online contribution. Attracting new members to online communities and encouraging substantive engagement are open and compelling problems at the intersection of multiple disciplines. My inquiry into social psychology theories as well as empirical studies related to community participation helped me identify a link between effective contribution and the psychological construct of self-efficacy [26, 215, 516, 543]. In addition, inquiry into the psycholinguistics literature informed me about text analysis methods I could apply in order to measure various personal attributes that the social psychology theories suggested as relevant to self-efficacy. This groundwork moti-
vated me to continue on a domain-driven approach to assessing and designing around this personal trait.

In essence, my analytic plan was to model self-efficacy using linguistic features of an individual’s passively mined text-based data. I therefore constructed various semantic, syntactic, and stylistic features intended to represent salient characteristics of self-efficacy according to my gathered domain knowledge (e.g., to operationalize attributes such as self-confidence, self-regulation, executive skill, critical thinking, and social competency) [27, 211].

I then collected data by acquiring two existing experimental datasets that were well suited to the task, would let me explore my domain-driven approach relatively cheaply, and would also help me assess its generalizability across more than one social context. In each experiment, the participants had contributed comments in an ad-hoc online community environment modeled after RegulationRoom [321], an established website designed to facilitate the contribution of feedback about proposed regulatory policies. For both datasets, a variety of participant information was also available, including self-efficacy measured using a validated survey [87].

Before moving onto the assessment phase of the framework (i.e., attempting to assess participants’ self-efficacy from their text-based data), I additionally used domain knowledge to verify a key assumption: that high self-efficacy actually translates into “effective” contributing behavior. Here, I used a definition of domain knowledge that included organizational theories, experimental evidence (e.g., about predictors of answer quality in Question & Answer sites), and publicly-posted community guidelines (e.g., Wikipedia policies) in order to define generally accepted metrics of effective contribution: comment
volume, length, and quality. Then comparing participants with strong versus weak self-efficacy (where the threshold was informed by prior literature [394]), I confirmed that individuals with strong self-efficacy did post statistically significantly more, longer, and higher quality comments.

Next, for both datasets, I compared individuals with strong versus weak self-efficacy and found statistically significant differences for nearly all my linguistic features. I additionally examined whether these features could predict self-efficacy, treating the task as a binary classification problem using the same strong versus weak self-efficacy groups. Using 10-fold cross-validation with a C4.5 decision tree algorithm, I found that my linguistic features were able to predict self-efficacy with reasonably good accuracy, precision, and recall and far outperformed a baseline model that returned the majority class.

I could have taken a data-driven approach to analysis, for instance by including numerous additional features (e.g., n-gram frequencies) and then using feature selection or other statistical techniques to prune. Instead, I focused on a grounded operationalization of psychological traits, constructing features that domain knowledge suggested would be more relevant to the task from the start.

To evaluate whether this domain-driven approach actually boosted predictive power, I performed the same binary self-efficacy classification task again, this time using more data-driven feature sets available from three recent relevant studies [186, 307, 544]. In fact, I found that my domain-driven set delivered the best performance, as it incorporated informative features that were not captured by these other sets’ dictionary-based categories. My intention is not to imply these studies are methodologically arbitrary but rather that they are representative of data-driven approaches that apply off-the-shelf text ana-
lytics instead of developing custom features to represent constructs of interest. Encouragingly, these results supported the idea that a computational strategy rooted in domain knowledge can enhance a phase of assessment (in a domain outside health).

Lastly, I was able to identify several ways these findings could help cultivate more fruitful communities through a domain-aware design process. For example, community recruitment could be enhanced by targeting self-efficacious individuals more likely to be willing to engage and do so in valued ways. Or, an intelligent task routing approach [110] that leverages self-efficacy profiling could improve the chances that contribution tasks would be funneled to individuals more likely to be willing and able to do them. Such an approach might even help to iteratively foster self-efficacy, given that the personal fulfillment and social recognition of completing tasks could provide members with mastery experiences, which is an effective way to develop a stronger sense of self-efficacy [27]. Finally, the increased understanding this work contributed about the relationship between self-efficacy and online contributing behavior could inform the design of moderator protocols to better direct attention or personalize engagements with users in ways suitable to their psychological characteristics.

Overall, my work indicated the generalizability of the domain-driven framework to online contribution. In particular, by applying the framework in this context, I was able to show how a suite of domain-driven features has useful prediction power, including compared to commonly used feature construction strategies that apply off-the-shelf dictionary-based methods without much eye to theory — bolstering the notion that a domain-driven approach may genuinely enhance assessment performance and, in turn, design opportunities.
6.3.3 Open Challenges of Data, Analysis, and Design

In addition to broadening PHI beyond a single-user model and expanding the framework to applications beyond health, a number of other open challenges exist that warrant further investigation. I focus on three issues that each relate to one of the three components of PHI technology: data concerns, analysis considerations, and possibilities for novel forms of feedback that engage users with PHI systems in new and meaningful ways.

Entire dissertations could be devoted to various facets of these topics. Nevertheless, while the following paragraphs cannot fully address all the myriad considerations, I believe there is value in at least recognizing these as important talking points, articulating their links to my particular dissertation, and offering concrete directions for the HCI community to pursue going forward.

Data Management — Ownership, Privacy, and Security

As increasingly diverse and granular personal data streams continue to be generated, either knowingly by individuals interested in tracking that information or as a passive byproduct of one’s interactions with technology, it is important to contemplate a number of concerns related to the responsible management of that data — especially considering health data is highly personal and potentially sensitive, stigmatic, and exploitable.

For instance, a move toward personal informatics paradigms that better support collective engagement, as described earlier in Section 6.3.1, will throw issues of data ownership, privacy, and access into sharp relief. In addition, data collection that is passive, streaming, and “always on” can hinder individuals’
ability to practice intentional self-disclosure and raises issues of privacy rights and data curation. An important step is determining the various motives, expectations, and concerns users would have regarding what data would be acceptable for systems to handle and how. There are numerous opportunities for research to examine ways to go about establishing legal standards and user-friendly terms of service; increasing systems’ transparency in how data is collected, stored, protected, brokered, or otherwise managed; and more deliberately providing usable mechanisms for opting in or out of certain aspects of digital health services. In pursuing these directions, some scholars have suggested that personal health data could be reimagined as a child, treating data with the same rights and responsibilities as would be given to a vulnerable, precious loved one [368].

The smartphone app I have been involved in developing to help people manage bipolar disorder (MoodRhythm, introduced in Section 5.4) builds in privacy-preserving mechanisms in several ways. For example, the system does not record audio but rather processes it in real-time on the smartphone in order to only extract and store features (e.g., spectral content, loudness) that are useful for detecting the presence of a human voice but insufficient to reconstruct speech content [537]. Using these privacy-sensitive audio features and probabilistic inference techniques, MoodRhythm is able to estimate whether a user is engaging in a healthy level of social interaction. In addition, accelerometer data is used to classify and securely store a user’s activity state as a binary active versus sedentary. Attempting to model specific activities would be more privacy invasive and computationally expensive — yet not necessarily more useful in detecting symptoms than MoodRhythm’s high-level, more privacy-preserving representation. In fact, a domain-driven approach that understands the specific
data that must be targeted and at what minimum level of granularity necessary to support management might better protect users’ privacy, be more well-received by users, and more easily comply with future standards.

The Elusiveness of Causality

When analyzing personal data or behaviors to assess health, observed relationships can be complex and multifaceted. In particular, I encountered an issue of elusive causality throughout my case study on sleep, performance, and emotional wellness. In Experiments 1 and 2 (described in Chapter 4), poor sleep could manifest in technology use, technology use could result in poor sleep, and both could have been modulating or modulated by a number of other factors. In many participants, I observed what seemed to be a cycle of disruption wherein they would get insufficient sleep, cyberloaf the next day due to problems with attention, and ultimately again lose sleep that night by staying up to compensate for such unproductive time. Observed links between sleep and mood were similarly difficult to unpack in those studies, as negative mood may both reflect and cause poor sleep, and both negative mood and poor sleep may be indicative of other underlying factors such as depression or stress.

In my studies on bipolar disorder, I ran into a similar chicken-and-egg question between technology use and condition symptoms. Many individuals discussed their uncertainty about whether particular patterns of use would generally trigger and fuel bipolar episodes or whether being manic or depressed would lead to distinct levels and types of technology use (e.g., “I also start to notice that I lose more and more sleep because I’ve been up late, usually online. Is losing sleep triggering episodes, or is the episode triggering my staying up late?”). I
also found that this interlaced relationship between usage and symptoms could become a positive feedback loop. For instance, somewhat elevated energy and mood levels could lead an individual to initiate conversations on a dating site or play games longer than usual — behaviors that would then increase stimulation, heighten energy and mood further, and eventually spiral into a full-blown manic episode. As another example I commonly observed, an individual experiencing stress would lose sleep as a result, increasingly use technology during periods of this insomnia, be more susceptible to negative posts online, experience stress yet again, and eventually fall into a phase of depression.

While it was beyond the scope of this dissertation to fully disentangle such intricacies, they are important to look for and attempt to unravel during PHI analyses, so as to avoid making mistaken causal assumptions as well as to determine what behaviors to actually target for intervention.

**Novel Feedback Formats**

In Section 2.2.3, I identified the ways in which PHI systems commonly convey feedback to individuals (e.g., using text, charts, visual metaphors, sounds, or vibrations). Being able to understand and derive value from such information requires a new form of literacy on the part of the end-user as well as new communication strategies on the part of PHI system developers. Moving beyond conventional approaches to data visualization may be particularly beneficial in some contexts, such as PHI tools for vulnerable populations with special requirements. For instance, while feedback can support healthy self-awareness, I found that many individuals with bipolar disorder struggle with traditional visualizations, which sometimes lead to hyper self-scrutiny, unrealistic normative
expectations of health or identity, or a distressing clash between the smoothed curves often emphasized by standard graphs and the sharply erratic fluctuations they associated with their own moods.

Recently, HCI researchers have begun exploring novel ways to engage users in their health information, for instance through storytelling experiences [179, 487] or tools that allow users to build custom visualizations themselves [21]. In my research, I am particularly excited about personal informatics games: tools that provide gameful and playful approaches to data capture, self-reflection, and behavioral intervention [357]. As an example, I am developing a game called Stress Fighter that incorporates biofeedback passively collected through off-the-shelf wearables in order to deliver stress-relief interventions. Gameplay dynamics are based on the classic arcade game Street Fighter, with attributes of the opponent boss character corresponding to the player’s sensed stress levels that day. Encouraging full body movement while experiencing the intervention, Stress Fighter continues to capture and incorporate real-time biofeedback during gameplay since physical exertion can also help tackle stress.

My next step is exploring how games can be used as part of cognitive or psychological therapy, inspired by games like EyeSpy [119], where searching for the approving face in a crowd of frowns is used to help recondition the mindsets of people with low self-esteem, or Play Attention [467], where neurofeedback is used to control game elements and improve ADHD symptoms. In particular, I am interested in the potential role of video games in nonpharmacological treatments for managing mental illnesses like bipolar disorder, for instance to help combat periods of depression or act as a safe outlet during manic episodes.
6.4 Concluding Remarks

The rapid evolution and dissemination of personal technologies have created unprecedented opportunities for enhancing health on a broad scale. This dissertation contributes to a growing area of HCI research aimed at realizing this potential by developing systems that capture personal data, use this data to assess health, and provide tailored feedback that empowers self-management.

Specifically, I have shown the value of a domain-driven approach that draws on diverse sources (e.g., scientific literature, empirical evidence, practitioner expertise, and user perspectives) in order to build a solid foundation on which development decisions can be made. I believe that an approach grounded in such domain knowledge can better target significant health determinants for assessment and intervention, provide designers with a deep understanding and empathy for the role of technology in a given health context, and, in my experience, lead to individuals’ downright enthusiasm in using these technologies.

Integral to this framework is a consideration for the context of application and the needs of users. In light of the tensions and open challenges I have discussed, I would like to underscore the value in taking a “do no harm” approach that carefully considers the promises and pitfalls of technology-based solutions, which have the potential to positively contribute to health self-management in meaningful ways when responsibly and compassionately developed.

Having only scratched the surface of opportunities for PHI tools, I see a vast design space to continue exploring going forward. I hope this dissertation provides a roadmap that helps researchers traverse a development path toward technologies that support effective, safe, and empowering experiences.


[30] Larissa K Barber and David C Munz. Consistent-sufficient sleep predicts


[40] Victoria Bellotti, Bo Begole, Ed H Chi, Nicolas Ducheneaut, Ji Fang, Ellen


[82] Centers for Disease Control and Prevention. The power of prevention:


[100] Sheldon Cohen, William J Doyle, Cuneyt M Alper, Denise Janicki-Deverts,


[110] Dan Cosley, Dan Frankowski, Loren Terveen, and John Riedl. SuggestBot:
using intelligent task routing to help people find work in Wikipedia. In
Proceedings of the 12th international conference on Intelligent user interfaces,

[111] Paul T Costa and Robert R McCrae. Revised Neo Personality Inventory
(NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Psychological Assess-


[113] Stephanie J Crowley, Christine Acebo, and Mary A Carskadon. Sleep,
circadian rhythms, and delayed phase in adolescence. Sleep Medicine,

[114] Mihaly Csikszentmihalyi and Reed Larson. Validity and reliability of the
experience-sampling method. In Flow and the foundations of positive psy-

[115] Giuseppe Curcio, Michele Ferrara, and Luigi De Gennaro. Sleep loss,
learning capacity and academic performance. Sleep Medicine Reviews,

[116] Charles A Czeisler, Jeanne F Duffy, Theresa L Shanahan, Emery N Brown,
Jude F Mitchell, David W Rimmer, Joseph M Ronda, Edward J Silva,
James S Allan, Jonathan S Emens, Derk-Jan Dijk, and Richard E Kronauer.
Stability, precision, and near-24-hour period of the human circadian pace-

[117] Serge Daan and Martha Merrow. External time–internal time. Journal of

[118] James M Dabbs, Rebecca Strong, and Rhonda Milun. Exploring the
mind of testosterone: A beeper study. Journal of Research in Personality,

[119] Stéphane D Dandeneau and Mark W Baldwin. The inhibition of socially
rejecting information among people with high versus low self-esteem:
The role of attentional bias and the effects of bias reduction training. Journal

[120] Munmun De Choudhury. Opportunities of social media in health and


[130] Nancy L Digdon and Andrew J Howell. College students who have an


[160] Dorothee Fischer, Céline Vetter, Christoph Oberlinner, Sven Wegener, and Till Roenneberg. A unique, fast-forwards rotating schedule with 12-h long...


[221] Maree L Inder, Marie T Crowe, Stephanie Moor, Suzanne E Luty, Janet D


[279] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim,


[290] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. Fish’n’steps: Encouraging physical activity with an inter-


[337] David C Mohr, Michelle Nicole Burns, Stephen M Schueller, Gregory Clarke, and Michael Klinkman. Behavioral intervention technologies: ev-


[347] Sai T Moturu, Inas Khayal, Nadav Aharony, Wei Pan, and Alex Sandy Pentland. Using social sensing to understand the links between sleep, mood, and sociability. In Privacy, Security, Risk and Trust (PASSAT) and


[355] Elizabeth L Murnane and Scott Counts. Unraveling abstinence and re-


[500] Wayne F Velicer, Colleen A Redding, Xiaowu Sun, and James O


[511] Ryan J Walker, Zachary D Kribs, Andrew N Christopher, Oren R Shewach,


