Goal-Centered Personal Informatics Tools

Yasaman Sadat Sefidgar

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:
James Fogarty, Chair
Sean Munson
Jeffrey Heer

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Goal-Centered Personal Informatics Tools

Yasaman Sadat Sefidgar

Chair of the Supervisory Committee:
James Fogarty
Department of Computer Science & Engineering

Personal data holds considerable potential for improving health and well-being: it can enable health understanding, management, and behavior change. However and despite the increasing abundance of personal data, a myriad of challenges come in the way of realizing this potential. Available systems often force recording data that is poorly connected to care needs and offer no or limited means of adjusting recording and aligning the data to the needs. End-users are on their own in working with the data, even though they frequently lack time or skills to implement their own data collection and analysis workflows. Difficulties linking data to effective actions further limit care opportunities. My dissertation demonstrates that health informatics tools with explicit representation of an individual's goals can help overcome these challenges in collecting, analyzing, interpreting, and acting on personal data. It does so through three systems: In MigraineTracker, I show tools which elicit an individual's goals can support highly personalized tracking (e.g., what, when, and how to track) and sense-making (e.g., what data to present and how). In Analyticons, I introduce constraint-embedded goal-aligned visual objects that scaffold end-user analysis of heterogeneous personal data. In WoNoB, I leverage goal pursuit techniques to enable reflection on personal data toward effective actions, despite changing contexts and competing needs. Together, my work establishes goal-directed design as a principled approach to giving individuals the agency to control systems and align them to their needs, while complementing their efforts with relevant computational and theoretical scaffolding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

PhD is rarely a solo activity. I was fortunate to have enjoyed collaborations that profoundly enriched the PhD and research experience. First and foremost, my advisor, James Fogarty, helped me grow as a researcher. He listened to my thoughts and guided me in refining them into fine ideas. Research discussions with James were practices of developing a keen eye for new and fundamental findings. James's input also taught me to write with more clarity. I am also immensely grateful to have enjoyed these opportunities in an environment without undue pressures and distractions.

Several other people mentored me throughout PhD: Sean Munson provided actionable and thought-provoking feedback on my work; I was often amazed by how these input furthered my understanding and thinking. Jeff Heer brought new levels of rigor to discussions and the research direction. Jina Suh taught me to be independent and to do rigorous work at speed. Paula Nurius patiently guided me in working through unfamiliar areas of psychology. Andrea Hartzler brought fresh perspectives that not only improved this manuscript but also highlighted multiple exciting future research.

My work became possible through generous support of our medical collaborators Natalia Murinova and Allison Cole. I am also truly grateful to the clinicians of MigraineTracker study who joined with a deep commitment to their patients. The study was extensive and happened separate from the clinical practice, which was already overburdened during COVID-19 pandemic. I am similarly grateful to MigraineTracker study patients who joined the study despite the burdens of their condition. I hope my work informs solutions that ease their burdens.

Fogies, old and new, were an important part of my PhD. I was fortunate to pick up on the great work Jessie Schroeder had started on goal-directed tracking. Jessie kindly answered my many

questions early on and offered suggestions and insights throughout the work. She is a key part of the success of MigraineTracker. Ravi Karkar and Alex Okeson helped me feel welcomed as I joined Fogies. Richard Li and I ensured Fogies meetings happened every week even with just the two of us. I hope we continue the lively conversations of those meeting moving forward. Mingyuan Zhong's tips and tricks on dealing with LLMs saved me a lot of time and headache. Anant Mittal's enthusiasm for system development was contagious and his suggestions for libraries and tools sped up Analyticons implementation. Jesse Martinez took the extra time of sitting with me and going over my early job presentations. Shaan Chopra's tips on qualitative analysis saved me time and effort. I am in awe by Tae Jones's ability for planning ahead for different situations. I had a lot of fun during the conferences I attended with Shaan and Tae and I hope to continue that tradition, along with other fun we had together. Shaan, Tae, and Aleyah Lewis also brought much needed energy to Fogies post COVID.

There were many other people who influenced my PhD experience. During their time at UW CSE, Sarah Chasins, Dominik Moritz, and Eunice Jun mentored me in my early PhD years. Being part of the large UW CSE community, I was fortunate to discuss research with many peers while having fun, e.g., in playing badminton with Rashmi Mudduluru, Alex Mariakakis, Edward Wang, Ruotong Wang, and Kevin Fang, in serving on departmental committees or advocating for students with Maarten Sap, Nicasia Beebe-Wang, Avery Mack, Miranda Wei, Pratyush Patel, Marissa Radensky, Naveena Karusala, and in hallway chats with Ashish Sharma, Jason Hoffman, Dhruv Jain, and Venkatesh Potluri. Dedicated staff at UW CSE were a key support system: Elise DeGoede, Joe Eckert, Chiemi Yamaoka, Andrei Stabrovski, Amy Ness, Jin Xu, Kellus Stone, and many others I did not meet but who helped me with my various requests. I had an amazing time at Microsoft Research's HUE team under Mary Czerwinski and I am immensely grateful for her continued support and advice.

My parents gave me strength with a mix of encouragement and pragmatism during the PhD journey. My conversations with my brother, inspired potential connections between my work

and other computer science topics including software design approaches and concept design. In addition to emotional support, my husband was a sounding board as I was brainstorming thoughts and solutions to technical challenges. He sat through my many practice presentations and is perhaps as familiar with my work as my advisor if not more. He also produced several images of this thesis based on my sketches. I delegated part of the development of Analyticons to my brother (data table interface) and husband (visual graphs).

DEDICATION

to the educators who believed in me

Chapter 1

INTRODUCTION

Over half of adults in the United States suffer from at least one chronic condition and about one third are affected by two or more [18], a rate similar to that of worldwide prevalence [84]. Given the globally aging population, these numbers are expected to grow, along with the social and economic costs [83]. These trends underscore the **need for prevention, monitoring, and management of chronic conditions**. The longitudinal measurements of symptoms, treatments, contributors, and well-being behaviors are critical to these needs [71]. However, the measurements **can rarely happen in clinical settings** due to the nature of the underlying factors. For example, symptoms can be intermittent and absent at the time of clinical testing. Behaviors and other factors unfold in daily life and cannot be observed in the confines of a controlled time and space.

Personal health and well-being data is a promising and complementary alternative to limited and often unavailable clinical data for long-term and chronic condition care. It can capture behaviors, symptoms, and other influencing factors as they develop over time and does so with increasing ease. People commonly carry smart phones and wearables and can record a multitude of information at the palm of their hands, both actively-reported and passively-sensed. Encouragingly, there is evidence for the benefits of the collection and use of personal data in health contexts. It can improve awareness and understanding of health [61], motivate and support behavior change [1], enhance health management and communication [123, 36], and increase engagement and agency across aspects of care, including diagnosis and planning [35, 160].

However, **the distinct potential of personal data is yet to be fully realized**. On one hand, personal data-related activities are challenging. It is non-trivial to decide what data to collect, when and how or in what ways to make adjustments over time. As a result, individuals

commonly are over-burdened by data collection that is largely irrelevant to their needs [32]. Analyzing and interpreting data in connection to one's care is also demanding. These activities require varied expertise, including data analytics and medicine, and overwhelm both patients and their clinicians [36, 160]. On the other hand, we lack effective tools to perform these activities. Incorporating personal data in complex and multi-faceted settings of health and well-being is often limited by rigid systems that frequently force recording data that is poorly connected to care needs and offer no or inadequate means of adjusting recording and aligning the data to the needs [76]. End-users are largely on their own in working with the data, even though they frequently lack time or skills to implement their own data collection and analysis workflows [31, 44]. Similar difficulties exist in linking data to effective actions in the absence of need-aligned support [29]. The recurring theme in these systems is that designers and developers assume certain needs and embed them in designs that inevitably fail to meet the diversity of needs people bring to the use of the systems. Hence and despite recent advancements that make it easier for people to get more data in kind and quantity, personal data technologies are yet to be truely human-centered.

In a shift from data-centered design, Schroeder et al. [173] proposed *goal-directed design* for personal data technologies: **designs should elicit and incorporate expressions of an individual's goals to then scaffold** the process of defining what, when, and how to obtain and use personal data for those goals. In their work, Schroeder et al. [173] also offered evidence for the benefits of this approach in addressing some of the difficulties of the planning stage of health tracking. **Whether and how effective this approach is** for other personal data activities, including collection, analysis, interpretation, and action, remain underexplored. Moreover, **little is known about considerations and requirements** for applying it across these activities.

My thesis addresses this gap through the design and deployment of systems, architectures, and techniques for leveraging goals. Through longitudinal deployment of MigraineTracker in the wild, I show systems designed around explicit goal expressions support diverse and evolving needs. I also identify key functionality to have when applying goal-directed design: alignment, evolution, and reflection [174]. I present WoNoB ('wanna be',

an acronym for work-nonwork balance) and Analyticons ('analytic' + 'icon') as illustrations of different approaches to building such functionality. WoNoB leverages self-regulation techniques of goal-pursuit for data review and reflection to help individuals identify opportunities for action and carry them out despite changing and competing needs [175], whereas Analyticons demonstrates an architecture for integration and analysis of heterogeneous personal data aligned with an individual's varied and evolving needs [176]. Taken together, my work demonstrates the following **thesis statement**:

Personal health informatics systems with explicit representations of an individual's goals can scaffold planning, collecting, analyzing, interpreting, and acting upon personal data.

1.1 Thesis Overview

I substantiate the claim of this thesis statement in three chapters. These follow Chapter 2, where I position the work in the literature on health and human-computer interactions (HCI), and present background knowledge on health conditions and psychology theories relevant to the work.

In Chapter 3, I present the deployment of MigraineTracker, the first major realization of the goal-directed design of personal informatics systems. My work on MigraineTracker built on Schroederet al. 's formative work and earlier app development. Briefly, MigraineTracker elicits goals (e.g., 'improving symptoms', and 'learning migraine frequency'), then uses them to guide end-to-end support for data collection (e.g., by recommending what to collect and how), presentation (e.g., via goal-aligned summaries and visualizations), and review (e.g., through goal-directed reflection). I deployed MigraineTracker in a 12+ month study with migraine patients and their clinicians, collecting data and insights in over 100 interviews. With this study, I showed personal data tools that elicit an individual's goals can support highly personalized tracking (in terms of what, when, and how to track) and sense-making (in terms of what data to present and how). Using the MigraineTracker, individuals tracked data for multiple and evolving goals, made adjustments when needed, and eventually better understood and managed their condition. I additionally contributed to the theoretical models of personal tracking (e.g., the

importance of accounting for multiple goals and their inter-relations in personal models) and highlighted key requirements for goal-directed system design. These include providing mechanisms for alignment and realignment of inter-related goals that evolve in relation to a reflective process where people form hypotheses, examine them, and refine different aspects of tracking. This work was originally published in the *ACM's CHI Conference on Human Factors in Computing Systems 2024* and won a Best Paper Award [174].

In Chapter 4, I explore a different approach to supporting goal-based designs with a focus on reflection in WoNoB. This system uses Implementation Intention [74] and Mental Contrasting [139], self-regulation techniques of goal-pursuit, to scaffold reflection process. Implementation intentions (IIs) are if-then plans that help individuals decide what to do to achieve their goals, where, and how and then act upon it. For example, 'whenever I use the restroom, I'll drink water' is an II plan for staying hydrated. Mental contrasting (MC) asks people to think about a desired state against their current reality and identify obstacles in reaching the desired state. When combined, MC provides a direction for action (overcoming obstacles) and IIs are formed to take actions. WoNoB walks individuals through a set of steps in examining data to find obstacles and opportunities for addressing them. I evaluated WoNoB in a three-week between-participant study with 43 information workers who used the system for improving work-nonwork balance in their day to day life. I showed implementation intention and mental contrasting (IIMC) can structure reflection on personal data, increase individuals' awareness, and allow them to more effectively find opportunities to act toward their goals despite disruptions and changing circumstances. Interacting with data without IIMC undermined behavior change. These findings highlight the importance of explicit scaffolding techniques for goal-oriented reflection.

In Chapter 5, I introduce Analyticons, an architecture that poses data analytics as assumption satisfaction and adjustments over concepts to handle the complexity of personal data, which is often heterogeneous with large variations between and within individuals. The heterogeneity is particularly pronounced in goal-directed data collection, where each individual may record different information, record the same information differently, or change their

recording over time. Analyticons architecture is based on constraint-embedded visual objects that are recommended given an individual's goal (e.g., a histogram of reported length of migraines is recommended for the goal of learning 'how long do migraines last?'). Selection of each object poses a set of assumptions over the concepts underlying the goal. The system then guides the individual in leveraging their unique knowledge of the data and its semantics to meet the assumptions. In the example above, the histogram requires ordinal values of duration concept. This assumption is used to scaffold a set of steps that allows individuals to identify relevant data and intuitively express how they want it transformed to ordinal levels. I showed core needs in managing chronic conditions such as migraines or irritable bowel syndrome can be expressed with Analyticons and illustrated different between and within variations that the system successfully handles. This work is an instance of new models of interaction to better support alignment and evolution in the integration and analysis of personal data around goals.

Following these chapters and in Chapter 6, I reflect on goal-centered design of personal informatics tools, where control over system functionality shifts from designers to users via different alignment mechanisms and as a result improves planning, collecting, analyzing, interpreting, and acting on personal data. I also point to the connection of goal-directed design to malleable interfaces and conclude by summarizing the contributions of my work to the design of personal informatics tools.

Chapter 2
BACKGROUND

In this chapter I review related work to highlight the importance of the problem considered in this thesis, and to position the arguments and proposed solutions. I first review work that shows the promise of personal data in the health context as well as the challenges impeding this promise, with a focus on chronic condition management (Section 2.1). I also introduce models of personal informatics as the theoretical grounding that guides my approach to where in the design space to explore and how to understand the experiences of self-tracking (Section 2.2). I then review related work on two specific subspaces I explored: reflection on personal data (Section 2.3) and personal data analysis (Section 2.4). I finally present relevant background on specific contexts I studied and why (i.e., migraine and work-nonwork balance in Sections 2.5 and 2.6) as well as the techniques I leveraged in creating some of the solutions (i.e., implementation intention and mental contrasting, Section 2.7).

2.1 Self-Tracking for Chronic Conditions

Self-tracking has long been considered a strategy to improve care and self-management of chronic health conditions [71, 122, 44]. Research has examined self-tracking for conditions with relatively well-understood symptom-contributor relations (e.g., asthma [90], diabetes [64, 123, 38, 122, 98, 159], hypertension [25, 70, 79]), conditions with enigmatic and intermittent symptoms (e.g., irritable bowel syndrome [171, 95], migraine [170, 173], multiple sclerosis [10, 198], polycystic ovary syndrome [33]), or progressive conditions (e.g., Parkinson's disease[131]). Such research has shown self-tracking can improve care and self-management through identifying factors which contribute to symptoms [96, 95, 171], control of symptoms [10, 198], and more effective collaboration with clinicians [36, 127].

Prior research suggests key features for self-tracking tools to support, including goal expression [137, 29, 37, 173, 62], guided and collaborative reflection [29, 44], customization [145, 29, 44, 76], and continuous learning [137, 44]. Research has also established that inappropriate designs of self-tracking tools may nudge people toward unwanted behaviors (e.g., an emphasis on calorie tracking promoting unhealthy eating [42]) and restrict data exploration and reflection [31, 29]. Tools may also promote unsustainably burdensome tracking routines [134] and undirected data representations that overwhelm people without answering their questions [103]. A struggle to find value in self-tracking [32] can in turn lead to abandonment [39, 61, 108].

Despite calls for supporting goal expression, particularly for qualitative and subjective goals [61, 137], it remains uncommon in current tools [29]. A notable exception is Schroeder et al. [173]'s goal-directed self-tracking framework, which proposes designs where explicit goal expressions drive *what*, *when*, and *how* an individual tracks. The framework aims to enable custom data collection and a need for goal evolution, which is integral to long-term tracking [137]. Although prior research has focused on increasing the amount or diversity of data people can collect (e.g., [105]), flexible tracking is not by itself sufficient when people are unable to connect data to their core needs [137, 103]. Goal-directed self-tracking therefore combines flexible tracking with a principle of reduction and focus [134] to emphasize tracking *exactly* and *only* data supporting an individual's goals. My work studies this framework, highlights its benefits, and identifies the key requirements for realizing these benefits (MigraineTracker, Chapter 3). It also contributes solutions for addressing those requirements (WoNoB and Analyticons, Chapters 4-5).

2.2 Models of Personal Informatics

Personal informatics models provide a lens for designing self-tracking tools and for understanding people's experiences. I have leveraged both Li et al. [111]'s stage-based model and Epstein et al. [63]'s lived informatics model in design (e.g., MigraineTracker) and in deciding where in the design space to explore. Li et al. [111]'s model characterizes distinct stages of

preparation, collection, integration (including data preparation and analysis), reflection, and action, highlighting how later stages depend on earlier stages. Epstein et al. [63]'s model additionally highlights lapsing and resumption in everyday experiences with tracking [165], which is particularly important in long-term tracking as with chronic conditions. I also considered recommendations stipulated by Niess et al. [137]'s Tracker Goal Evolution Model, which contextualizes goals within the lived informatics model. It highlights that needs (e.g., 'feeling well') manifest in qualitative goals (e.g., 'losing weight') which are translated into quantitative goals (e.g., 'taking 12K steps'). These quantitative goals can be linked to data in self-tracking tools (e.g., 'step counts'). The model thus highlights the importance of considering qualitative goals and supporting their translation into quantitative goals as part of meaningful long-term engagement with tracking.

2.3 Reflection on Personal Data

Reflection on personal data is key to using it in supporting health and well-being. It is recognized as a major activity around personal data [111, 63]. Benefits of reflection have been demonstrated in supporting well-being in general (e.g., encouraging physical activity [109] and healthy eating [179], managing stress [87]) and in the management of chronic conditions (e.g., diabetes [121], irritable bowel syndrome [37, 171]). Reflection is a multi-layered meaning-making process where individuals describe their experiences (or representations of the experiences), explain the patterns, and examine relations among the patterns, potentially in relation to the broader context of the experience [67]. The key aspects of a reflective process include engagement with breakdowns in one's "current meaning perspectives" through an inquiry process and in a dialogue with one's data. This process then transforms one's initial state of understanding [11].

Given the central role and benefits of reflection in personal informatics, personal data tools often include features to support reflection [60]. This is often happening in the form of visual representation of data [7, 14, 30, 31, 88, 91]. Despite the attention, reflection remains challenging [161, 29]. It is often assumed that including features that *can* support reflection,

means reflection happens and produces health benefits (e.g., by informing health-related actions). Neither of these assumptions holds [189, 134], underscoring the importance of designing different forms of scaffolding to guide individuals in reviewing their data, and doing so directed at supporting health goals. Adding to the growing body of work to address this gap (e.g., [121, 159, 197]), my work in Chapter 4 presents an approach to scaffolding reflection in support of taking goal-aligned actions.

2.4 End-User Analytics for Personal Data

Data analytics is among critical personal data activities in models of personal informatics [111, 63]. As the process of preparing data (e.g., cleaning and combining sources) and transforming it (e.g., as metrics or visual representations), it is a prerequisite to sense-making and acting. Given it is closely linked to reflection, most available systems often offer it as a standard pre-defined pipeline supporting summary or visual information [60]. However, the generic processing features usually fall short in supporting individuals to answer specific questions [172]. Along with growing calls for goal-centered personal informatics [29, 44, 173], support for customized data analytics consistent with people's diverse health and well-being needs is important.

Developing tools for end-user data analytics is an active area of research with solutions to empower individuals in data cleaning [158], transformation [93], reshaping [192], exploration [178, 191], and visualization [146, 183]. With recent advances in natural language processing, producing data summaries and visualizations with text input is on the rise, although the focus is often on iterative exploration rather than data wrangling [26]. However, existing solutions make certain assumptions that limit their use for personal data analytics. Some assume standard semantics (e.g., no outliers) and layout (e.g., tidy) for data [146, 183]. With unmet assumptions, these solutions are either unusable or unreliable. The latter is concerning if processing details are hidden from users or are beyond their expertise to examine and verify (e.g., in case of off-the-shelf interfaces, such as ChatGPT). When enabling operations for non-standard data, existing solutions frequently assume individuals know what operations are

relevant and thus only support them in taking them (e.g., [93, 192]). My work in Chapter 5 explores analytics solutions that assume neither standard data nor analysis expertise.

2.5 Migraine as a Context for Examining Goal-Directed Self-Tracking

Migraine is a debilitating chronic condition that can result in reduced quality of life [181], occupational impairment [124], constrained social and family functioning [180], economic burden [65], and diminished emotional health [47]. There is high idiosyncrasy in migraine symptoms [147], which are often simultaneously affected by multiple and accumulating factors [104]. Managing migraine relies on medication and behavior changes to limit contributing factors and to encourage preventive and abortive measures [53]. However, there is high variability in response to medication [125] or behavior change [4]. These characteristics make self-tracking particularly useful for managing migraines [170].

Despite the potential, self-tracking in migraine is challenging as current tools are generally not well-aligned with an individual's evolving needs. Tools commonly force individuals to record irrelevant information (e.g., recording irrelevant contributing factors) or fail to support recording needed information. Individuals therefore struggle in preparing or in adjusting what they track and often fail to obtain useful information [170]. Addressing these challenges, Schroeder et al. [173] proposed goal-directed self-tracking, a design framework wherein explicit scaffolding for goal expression guides individuals to (1) track *exactly* and *only* the data they need and to (2) review data in the context of goals. Examining this approach with a paper prototype, they found improved tracking preparation and anticipated benefits for all stages of self-tracking. I expand this work with a functional prototype (MigraineTracker, Chapter 3) to examine whether and how goal expressions facilitate data collection, reflection, and action. I also consider the specifics of data in self-tracking of migraine and offer solutions to enable end-user analysis of this kind of data (Analyticons, Chapter 5).

2.6 Work-Nonwork Balance as a Context for Leveraging Personal Data

Well-being in the workplace is important for work engagement, productivity, and job satisfaction [163, 182]. A positive organizational culture that promotes employee well-being is associated with positive interpersonal relationships [6, 52], reduced absenteeism, and reductions in related costs [119, 41]. One of the key facets of well-being in the workplace is work-nonwork balance [194, 54, 168]. Delecta [48] defines work-nonwork balance as an individual's ability to meet their work and personal commitments, as well as other non-work responsibilities and activities [48]. The inability to strike the right balance between work and nonwork roles or demands can lead to over-commitment to fulfill the responsibilities of both [50], with detrimental effects on the overall well-being of workers [120, 196] due to the spillover of stress between work and life outside of work [80].

Work-nonwork balance sits along a continuum between integration (i.e., blurred boundaries) and segmentation (i.e., strong boundaries), based on the degree that work is kept separate from nonwork [5, 8, 23]. Although work-nonwork balance has been conceptualized as an individual preference, scholars argue that work-nonwork balance should not be portrayed as only a matter of individual choice. It is a socio-cultural phenomena that should rather be jointly considered with organizational policies, norms, and expectations [142, 69, 19]. For example, organizational policies such as flexible work arrangements have been shown to be effective in improving work-nonwork balance [17] and organizational culture has been shown to influence the utilization of work-nonwork balance programs [22]. Therefore, work-nonwork balance goals, while personal and idiosyncratic in nature, are affected by the surrounding work environments, such as work demands, culture, social dynamics, or flexibility.

In recent decades, the proliferation of information and communication technologies (ICTs) that enable virtual work, the use of personal technologies at work and vice-a-versa, and the always-on work culture have all contributed to an increased blurring of work-nonwork boundaries [126, 49]. Such blurring of boundaries has been further exacerbated by the mass transition to remote work during the pandemic [28, 153]. Unfortunately, the flexibility that is

afforded by ICTs has also made it more difficult to achieve work-nonwork balance [157, 40, 166]. Multiple surveys across sectors report that over half of Americans experience poor work-nonwork balance [187, 77]. Therefore, supporting work-nonwork balance in the digitized world has become a pressing need. It requires work on multiple fronts, from empowering individuals to take actions that better their work-nonwork balance, to getting at the social and cultural factors that lead to worsening work-nonwork balance. This motivates my focus on enabling individuals' actions to improve their work-nonwork balance within the bounds of external factors.

Workers are genuinely interested in improving their work-nonwork balance, but struggle to do so [78]. Realizing well-being goals is generally challenging [75], and is even more so in the context of work-nonwork balance because high demands and external pressure lead to competing goals and because people are particularly vulnerable to giving in to more pressing needs [141].

Literature in human-computer interaction (HCI), computer-supported cooperative work (CSCW), and ubiquitous computing (UbiComp) has shown extensive interest in understanding and supporting worker well-being [43, 2, 99, 164, 101]. For example, Das Swain et al. [46] studied how the routine-fit of employees with their workplace associated with employee functioning and well-being [46]. Rudnicka et al. [166] proposed a tool for promoting flexible social norms of break-taking at work for remote workers [166]. Cambo et al. [24] developed a mobile application, BreakSense, to promote mobility during work breaks, and examined how that impacted people's sense of completion and readiness to work [24]. Epsteinet al. [59] studied the relationship between work breaks and productivity through a self-reflection tool of summarizing breaks [59].

Adding to this research, I use personal data to support work-nonwork balance. In particular, I draw on self-regulation techniques (implementation intention and mental contrasting) to enable the effective use of personal data in improving work-nonwork balance (WoNoB, Chapter 4).

2.7 Implementation Intention and Mental Contrasting

Goal setting and realization are the primary predictors of health behavior change, with the latter explaining much of the variations in goal achievement [75]. However, goal realization poses challenges for many people; about half the people with intentions to engage in health behavior fail to do so as they cannot successfully translate their intentions to actions [144, 177]. There are three processes underlying this intention-action gap: viability, activation, and elaboration of intentions. Intention viability describes how abilities, resources, and opportunities available to a person influence whether they have control over certain behaviors. Intention activation is the process by which situational demands impact cognitive and motivational resources and thereby the salience and intensity of intentions. There are often conflicting goals and people are particularly inclined toward more enjoyable or pressing alternatives. They either deprioritize goals initially set in favor of those alternatives or forget about the initial goals altogether. Intention elaboration refers to the role that the clarity around actions and contextual factors plays in making goals feasible. Behavior change for health and well-being is often a complex sequence of actions. Failure in identifying these actions and the means to perform them undermines goal attainment [154].

Implementation intention (II) and mental contrasting (MC) are self-regulation techniques that support goal realization by mitigating the processes that underlie the intention-action gap [138]. While goals are specifications of a *want* ("I want to do/achieve X"), IIs are if-then statements that specify the *when*, *where*, and *how* of achieving a goal. They take the form "if situation Y happens, I will do [goal-directed] action Z" and connect situational cues to specific actions. For example, "If it is break time, I'll walk down to lobby" for the goal of "being physically active". The act of creating IIs mitigates the elaboration challenge as the deliberation on the relevant actions and their means is a part of the II plan creation process. By connecting specific situations to actions, the planning process becomes more focused on available resources, which helps mitigate viability challenges to the extent possible (e.g., by taking into account the constraints imposed by social and cultural factors). Forming this connection additionally supports maintaining goal activation levels as control of the response is delegated to the presence of the situation [154].

MC involves thinking about a desired future (e.g., "excelling in an exam") and juxtaposing it with the current reality that impedes achieving that future (e.g., "disorganized notes") [138]. Doing so increases motivation for action if there is a high expectation of achieving the desired future [138]. Heightened motivation subsequently supports goal pursuit [117]. The association between future and reality additionally focuses attention on dealing with reality and the instrumental means for it [139], which are important for addressing viability and elaboration challenges in goal attainment.

II and MC have successfully supported goal realization in a variety of domains such as health [45, 154], education [73, 12], or interpersonal problems [140, 74]. They provide greater benefits when combined [138]. MC complements II by providing a concrete process for forming if-then plans, where obstacles form the if-part and overcoming them forms the then-part. It thus mitigates the challenge of plan formation in applying II [154]. II, on the other hand, complements MC by facilitating the process of overcoming difficult obstacles through explicit planning [138]. The combination of II and MC (IIMC) is particularly well-suited for realizing in such contexts as work-nonwork balance given they can address challenges in viability (e.g., constraints in time, energy, or other resources available to the individual), activation (i.e., competing goals), and elaboration (e.g., changing situations) that arise in this context. In fact, IIMC has successfully supported the pursuit of the related goal of time-management [139]. Bringing these techniques to the use of personal data in work-nonwork balance improvements is thus a promising avenue for further explorations.

IIs are not new in HCI research on health and well-being. However, much of the past work is concerned with supporting the execution of II plans rather than making them. For example, Pinder et al. [152] explored how a context-aware smartphone app can support people by automatically detecting critical situations and reminding the user of the actions to take. Similarly, Bharmal et al. [16] explored the use of peripheral reminders to enhance the activation of goal-directed actions to increase physical activity. Both of these pieces of work [152, 16] focus on enhancing the link between situations and actions in II plans rather than supporting the plan creation itself. Dogangun et al. [51] modeled daily routines to automatically identify

and recommend timeslots or situations that can be used as critical conditions in IIs for physical activity; while this design addressed plan creation, it neither supported users in taking control of their time nor helped them identify obstacles – a critical activity in interpersonal contexts such as the workplace. In my work (WoNoB, Chapter 4), I leverage IIMC in reviewing personal data for creating behavior plans that empower individuals to take actions within their individual and organizational contexts with full agency.

Chapter 3

GOAL-DIRECTED SELF-TRACKING OF MIGRAINE

Management of chronic conditions often involves examining one's health and making adjustments in behavior or treatment [71], typically based on measures of symptoms, contributors, or treatments. Self-tracking of such measures is thus common in managing conditions such as migraine [147], where intermittent symptoms over time limit utility of clinical testing. However, self-tracking for chronic conditions like migraine is challenging, in part because existing tools embed assumptions about what people want to gain from tracking (i.e., their *goals* for tracking [170]). Deciding what to track, adjusting tracking over time, and using tracked data is thus poorly supported [173]. This gap in tracking support is particularly problematic given high idiosyncrasies among migraine patients and the complexity of potential contributors [170]. Schroeder et al. [173] thus proposed *goal-directed self-tracking* as a framework to address this gap. They suggested that designs centered around an individual's goals can support tracking *exactly* and *only* the data that individual needs. More concretely, tools designed within this framework aim to elicit goals and scaffold a process of defining *what*, *when*, and *how* to track toward those goals. Goal-directed tools can also use knowledge of individual goals in determining what data to present and how to better facilitate interpretation. In their work, Schroeder et al. [173] offered evidence that this approach improved preparation for tracking. However, their method did not allow them to examine if the improvements extended to data collection, reflection, and action.

In this chapter, I present *MigraineTracker*, an app designed and developed to understand the lived experience of goal-directed self-tracking and to examine whether and how tracking routines configured within a goal-directed tool can support patients in managing their migraines across stages of tracking [111, 63]. The study engaged 10 patients, each working with a clinician and

using the app for average of 12+ months, totaling 107 interview sessions. This chapter is based on the 50 sessions that focused on patient experience and contributes the following. I use the first person plural to describe the work and its contributions to acknowledge the important work by all collaborators involved in this multi-year project. While I led the completion of MigraineTracker development, design and preparation of the user study and its material, participant recruitment, study planning and coordination, study sessions, data analysis, and manuscript preparation this work was impossible without Carla Castillo, Shaan Chopra, Liwei Jiang, Tae Jones, Anant Mittal, Hyeyoung Ryu, Jessica Schroeder, Allison Cole, Natalia Murinova, Sean Munson, and James Fogarty, who contributed to different aspects of the work in various capacities.

- We demonstrate goal-directed data collection and reflection supports patients in (1) deciding what to track and how to align their tracking to their needs, (2) obtaining relevant and useful knowledge from tracking, (3) recognizing when and how to adjust their tracking, (4) feeling prepared to discuss their care with their clinicians, and (5) seeking expertise where they most need it.

- We provide evidence that goal-directed tracking led patients to further understand their condition and to feel they were better caring for themselves.

- We highlight the need for adapting personal informatics models to consider distinct and concurrent goals that are each at a different stage of tracking. We also articulate distinctions and relations among goals as an analytical lens for understanding needs and challenges in long-term self-tracking of chronic conditions.

I next present our design of MigraineTracker (Section 3.1) and describe our deployment study and analysis (Section 3.2). I then share key observations around goals, their evolution, and the benefits of goal expressions for patients (Section 3.3). I conclude by discussing implications for future research in personal informatics and the design of self-tracking tools (Section 3.4). This work was published in ACM CHI Conference on Human Factors in Computing Systems (CHI'24).

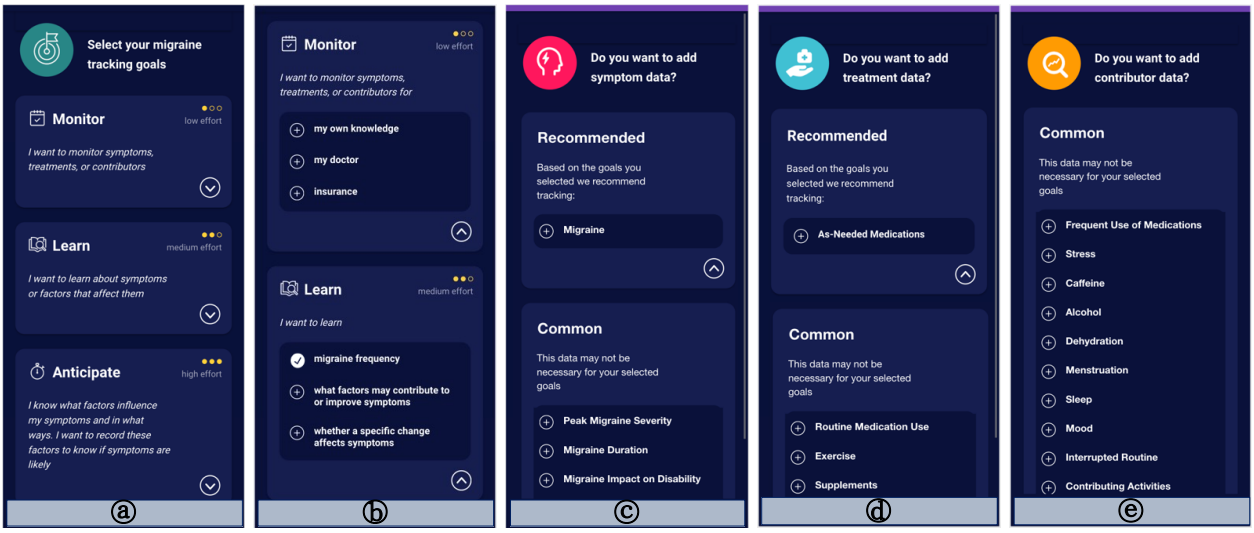


Figure 3.1: MigraineTracker configuration process. (a) Individuals can express goals in three categories: monitoring, learning, and anticipation. (b) There are sub-goals under each category. For example, monitoring for 'my own knowledge' or 'my doctor' and learning about 'migraine frequency' or 'what factors may contribute to or improve symptoms'. Recommended tracking items based on an individual's configured goals (e.g., learning about migraine frequency) appear under categories of (c) symptoms, (d) treatments, or (e) contributors.

3.1 MigraineTracker System

We built upon the formative work of [170, 173] to design and develop MigraineTracker in a human-centered design process. Design was iterative and involved cycles of paper prototyping, development, feedback from the research team (e.g., co-authors with lived experience and/or clinical expertise in migraine), and revision. There are three major components to MigraineTracker: configuration (Figure 3.1 and 3.2), data entry (Figure 3.3), and data review (Figure 3.4). Configuration is available at the onset of tracking and is modifiable afterwards. Data entry is defined by configuration and provides the interface individuals regularly use to record information. Review is available as a calendar visualization within the app and is complemented by more sophisticated summaries and visualizations outside the app. We provide additional details on each component below.

3.1.1 Tracking Configuration

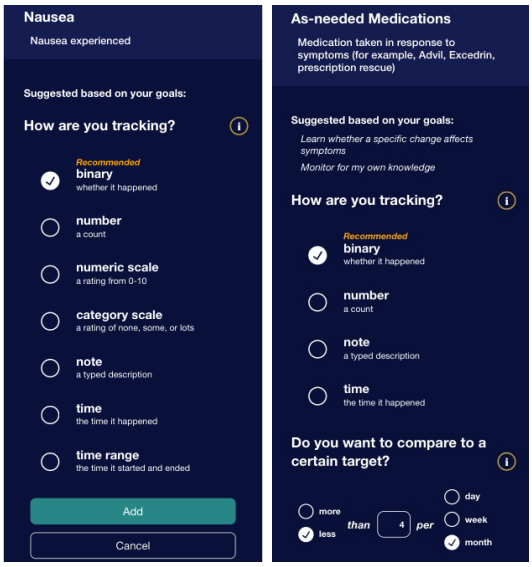


Figure 3.2: Configuration of tracking routine in MigraineTracker. The left shows nausea can be recorded in a number of ways (e.g., whether it happened–binary, its level–category scale). The right illustrates setting targets for as-needed medication (a count of less than four per month).

Goals are at the center of MigraineTracker configuration. As such, configuration starts with *goal expression* through selecting goals from three categories: monitoring, learning, and anticipation (Figure 3.1, a-b). Next is constructing a *tracking routine* (i.e., selecting *what* and *how* to track) in three categories: symptoms, treatments, and contributors (Figure 3.1, c-e). Items are recommended based on an individual's selected goals. There are also common items that may or may not be relevant to an individual. These items are separately listed to discourage tracking more than necessary. Custom items allow recording information that does not appear in the 'Recommended' or 'Common' lists. For example, MigraineTracker recommends recording 'Migraine' when an individual has a goal of learning about migraine frequency. It is also possible to record 'Peak Migraine Severity', which is commonly considered along with frequency. After selecting what to track, an individual decides *how* to record each selected item. For example, if they choose to record nausea, they can decide to record whether they experienced nausea or they can rate the levels of their nausea (Figure 3.2, left). There are recommendations for such data types based on an individual's goals. An individual configuring treatment and contributor items can also specify targets to get indications of behaviors relative to the targets. For example, setting a target for the dosage of as-needed medication (Figure 3.2, right) provides an indication of status relative to the target (Figure 3.3, b).

Configuration also supports reminders. The app offers up to two daily reminders and/or followup reminders. Followup reminders alert within a specified period of time (e.g., a day later), allowing individuals to initially report symptoms and then fill in other details at a later time when they have recovered.

3.1.2 Data Entry

With a configuration in place, individuals can record data for their selected items. Certain items (e.g., 'Migraine') appear on the landing page for quick entry (Figure 3.3, a). Others are listed by category (e.g., symptoms, treatments, contributors) so an individual can navigate to them as needed (Figure 3.3, c-d). MigraineTracker provides a per day data model where information is recorded against each calendar date. Although it is possible to support semi-automated tracking within the app, the version used in our study only supported manual entry. MigraineTracker also offers a lapsing feature, which pauses all notifications (e.g., for a vacation) with a configurable reminder to resume tracking at a specified date.

3.1.3 Data Review

A landing page calendar provides a simple view of tracked data (Figure 3.3, a). Migraine days appear with a bright pink background (e.g., Sep 03 in Figure 3.3, a), whereas days with other symptoms have a dark purple background (e.g., Sep 02 in Figure 3.3, a). Small colored dots indicate information has been tracked within a category. For example, a light orange dot indicates information was recorded under the 'Contributor' category (e.g., Aug 31 in Figure 3.3, a).

The research team also prepared static data summaries and visualizations, personalized according to tracking goals and preferences of each participant (Figure 3.4). These were prepared as needed throughout the study (e.g., when patients were meeting with the research team, when patients requested them for appointments with clinicians). These were not available at other times. We opted for this approach to enable iterative and exploratory preparation of a set of goal-based visualizations, which was not feasible within the app.

Figure 3.3: MigraineTracker data entry and review. (a) Calendar view with migraine days in bright pink and days with other symptoms in dark purple. Colored dots indicate tracking of information from that category. (b) Quick tracking items appear on the landing page, including their status relative to relevant targets (e.g., for as-needed medication). (c) Tracking items are organized under categories for symptoms, treatments, contributors, and other. (d) Opening a category presents configured tracking items.

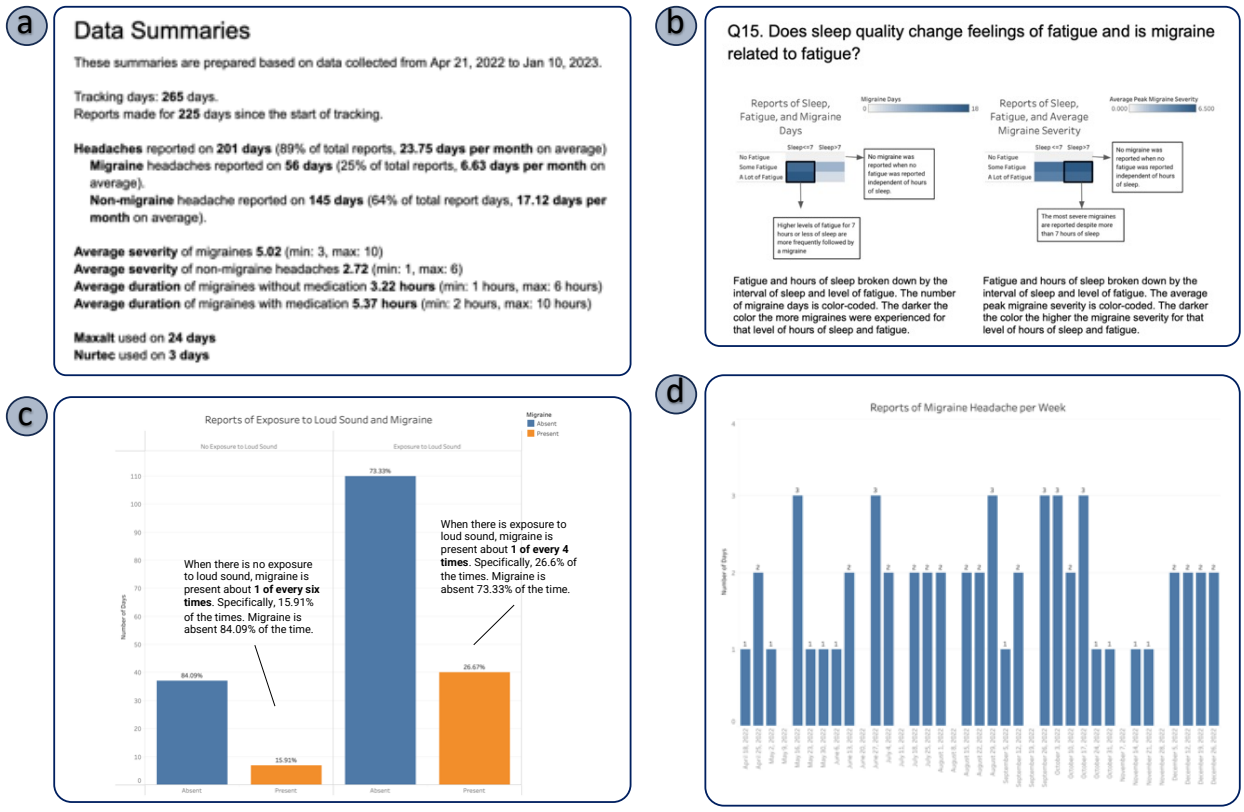


Figure 3.4: PT06's (a) data summary and sample visualizations: (b) inter-relations of sleep, fatigue, and migraine based on the occurrence and severity of migraines at different levels of fatigue and sleep, (c) frequency of the presence and absence of migraines with vs. without exposure to loud sound, (d) number of migraine days per week over time. Summaries and visualizations were separately prepared for each patient as needed throughout the study (e.g., when patients met with the research team or their clinician).

3.2 Deployment Study

We used MigraineTracker as a technology probe [89] in a deployment study to examine the lived experience of self-tracking with tools designed according to a goal-directed framework. This study builds on Schroeder et al. [173]'s investigation of whether patients can successfully use a tracking tool explicitly configured for their goals and further examines needs and considerations in designing goal-directed self-tracking tools for different stages of tracking [111, 63]. Examining goal-directed informatics in the context of migraine, our primary research questions are:

RQ1 How does tracking based in explicit expression of goals support patients in managing their migraine?

RQ2 How do patient goals and tracking change as they use MigraineTracker over time?

3.2.1 Recruitment and Participants

We advertised the study to migraine patients on mailing lists and via flyers, then reached out to their clinicians to join the study. If a patient's clinician was unavailable for the study, we offered to match the patient to a clinician already participating in the study. We also separately recruited clinicians through clinical collaborators and asked them to refer patients to the study. This approach of reaching clinicians through patients and vice versa increased our chances of recruitment during the pandemic. We asked clinicians to refer patients who experience migraine. Although we did not require a formal diagnosis of migraine, we also did not recruit patients who had a different specific diagnosis (e.g., cluster headaches). Patients who enrolled without clinician referral self-identified as experiencing migraines. Patients were in the United States, over 18 years old, and owned an Android or iOS phone to run MigraineTracker. Both headache specialists and primary care physicians were recruited, as both commonly work with migraine patients and not all patients have access to specialty care.

We initially recruited 17 migraine patients and five clinicians, of which 10 patients (eight women) and three clinicians (all women) completed the study (Table 3.1). We removed four patients we identified as inauthentic, an increasingly common challenge in remote research [162]. Three patients left the study after the initial interview and before starting tracking: one because their clinician did not join the study and two because their schedules changed. Two clinicians withdrew due to the demands of the ongoing pandemic. We observed both established patient-clinician pairs (9 pairs) and a newly formed pair. All participants were new to MigraineTracker (i.e., they had not participated in prior activities that informed the design of MigraineTracker).

Table 3.1: Patient demographics and length of study tracking. Gender was self-reported, consistent with recommended practices [169]. Length of study tracking is from the day they configured the app to their final day of recording. Patients either worked with their own clinician ('Y' under 'Established?') or a clinician they were matched with for the study. PR01 and PR02 were headache specialists and PR03 was a primary care clinician.

	Age	Self-Reported Gender	Study Tracking Days	Clinician	Established?
PT01	49-64	Woman	420	PR02	Yes
PT02	34-48	Woman	471	PR02	Yes
PT03	34-48	Woman	422	PR01	Yes
PT04	34-48	Man	420	PR01	Yes
PT05	18-33	Man	355	PR03	No
PT06	18-33	Woman	373	PR01	Yes
PT07	34-48	Woman	398	PR02	Yes
PT08	18-33	Woman	269	PR02	Yes
PT09	18-33	Woman	367	PR01	Yes
PT10	49-64	Woman	335	PR02	Yes

3.2.2 Procedure

Figure 3.5 provides an overview of our study protocol. After screening for inclusion, we interviewed patients about their self-tracking needs and prior experiences. We also introduced patients to goal-directed self-tracking and installed MigraineTracker on their phones. After demonstrating its basic functionality and use, we asked patients to think aloud as they configured their tracking routine.

After configuring the app, patients started tracking. After an average of two and half months of tracking (range: 27-153 days), they met their clinician, discussed their tracking setup, and made modifications if desired. This process ensured both patient and clinician goals were considered in tracking. In preparation for this session and to ease patient-clinician interactions, we briefly

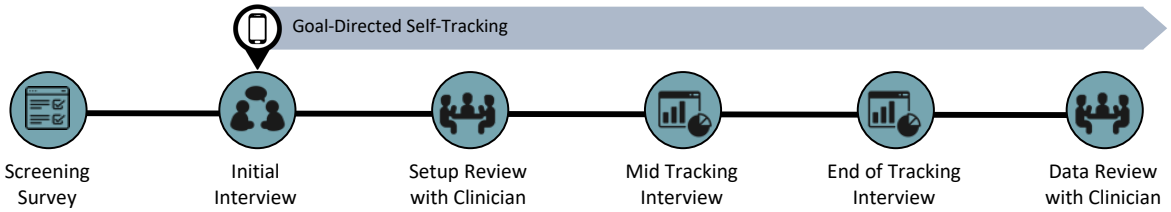


Figure 3.5: Plan for longitudinal study of migraine. We recruited patients who met the recruitment criteria in their screening responses. We learned about their needs and self-tracking experience in the initial interview and helped them configure MigraineTracker. Patients next reviewed their setup with their clinician. We learned about patient experience using the app in mid tracking and end of tracking sessions and obtained their feedback on goal-appropriate data summaries and visualizations. Patients then met with their clinician to review their data and material. Patients recorded information using MigraineTracker as soon as configuring it. Some patients continued tracking even after the second meeting with the clinician concluded their participation.

presented MigraineTracker to clinicians and created summaries of each patient's tracking setup. We were also available to answer questions as we observed these interactions. We did not, however, provide any formal training or guidance to patients or clinicians, in order to avoid overly influencing study results. We asked patients to track consistent with their goals and informed them if their tracking was inconsistent (e.g., if they had configured goals that required everyday tracking but only tracked when having symptoms). We documented these incidents as probe-surfaced needs and requirements for future support.

We met with each patient twice after their setup review with the clinician. The first meeting, a mid tracking interview, was scheduled an average of three months after the setup review (range: 46-144 days), when patients had collected a reasonable amount of data with respect to their goals. The second meeting, an end of tracking interview, was scheduled before a second meeting with the clinician. The main purpose of these sessions was to learn about patient day-to-day tracking experiences and any changes they made in their tracking. We also obtained feedback on goal-appropriate summaries and visualizations we had prepared based on patient and clinician comments in earlier sessions. This material was not available at other times unless patients requested it (e.g., for clinical visits outside the study). Following the end of tracking interview,

each patient met their clinician to review goal-appropriate material, interpret patient-generated data, and make decisions about patient care. We conducted a short follow-up interview with clinicians after each session with a patient. Patients and clinicians reviewed materials in the same static formats used in mid tracking and end of tracking sessions, sometimes updated according to patient feedback from those sessions.

Patients completed weekly check-in surveys throughout tracking to report their self-tracking experiences, difficulties, care status (e.g., any scheduled appointments with their clinician), and changes in their tracking goals or routine. We addressed any critical usability issues that were raised in these reports. In addition to weekly surveys, patients completed surveys about their experience each time they met with the clinician as a part of the study. If patients requested their data for a non-study visit with their clinician, we followed up on their experience during that clinical visit.

Patient tracking data was continuously synchronized to our database with daily backups. We shared an exported copy of the data with each patient at the end of the study. We could directly access and review data to prepare study material and to help with any technical difficulties. Our analysis and presentation of patient data was performed in collaboration with patients and according to their goals for that data.

We performed most initial interview sessions remotely due to pandemic constraints. We were later able to conduct most mid tracking and end of tracking sessions in-person while adhering to safety protocols. All meetings involving clinicians were in-person and subject to the same safety protocols. In-person patient sessions occurred on our university campus to ensure these feedback sessions were distinguished from clinical practice. Patient-clinician sessions generally occurred in the same clinic that a patient and their clinician typically met. This increased external validity and simplified logistics for patients and clinicians.

The initial, mid tracking, and end of tracking interviews took an average of about 90 minutes. Setup review and data review sessions lasted for an average of about 45 and 60 minutes respectively. We compensated patients at a rate of $10 for each 10 minutes of their time in sessions with the research team or their clinician. Patients also received $5 for each

weekly survey they completed and $10 for responding to surveys after meeting their clinician. Clinicians were compensated at the rate of $20 for each 10 minutes of their time, unless they chose not to be compensated. This study was reviewed and approved by our institutional review board.

3.2.3 Analysis

We analyzed data in multiple ways for different needs of the study. Early sessions (i.e., initial interview, setup review with clinicians, mid tracking interview) were analyzed to prepare for later sessions (e.g., for the end of tracking interview). For example, we analyzed the initial interview to understand patient goals and the mid tracking interview to understand whether and how patient goals evolved. The research team regularly met and discussed observations and our interpretations in preparing for upcoming sessions. At the conclusion of the study, we used interview transcripts, session notes, survey responses, tracking setup, and tracking data to summarize patient experiences in vignettes. Vignettes provided an overview of key observations in each patient's experience and familiarized the entire research team with each patient across study sessions. Lastly, we performed reflexive thematic analysis [20, 21], focused on the end of tracking interviews, to develop themes around patient experiences with goal-directed tracking. We decided to focus on end of tracking interview sessions for formal analysis as these provided the most comprehensive account of patient experiences, and the protocol for each end of tracking interview was informed by our analysis of multiple earlier sessions. In this process, we drew on our expectations and questions as personal informatics researchers and on our positions as designers of MigraineTracker within the goal-directed self-tracking framework. Another collaborator (Carla Castillo) and I used a combination of inductive and deductive coding to analyze end of tracking sessions. Deductive coding was informed by models of personal informatics (e.g., stages of tracking) and the goal-directed tracking framework (e.g., goal-centered configuration, goal evolution). Inductive coding occurred iteratively as we constructed new themes. Although we focused coding on the end of tracking interviews, Carla and I referred to other sessions, survey responses, and tracking data

as needed to support further understanding or provide key details. We double-coded four of the 10 sessions, compared analysis, and discussed themes throughout analysis. Each of us wrote memos as we reviewed data. The research team as a whole developed inductive codes through the coding process as well as discussions and memos about key observations. Initial themes were based on patterns in the data (e.g., 'new or refined goals', 'adjustments in data entry process') and were grouped to form higher level themes (e.g., 'distinct goal types', 'tracking models', 'alignment between goals and tracking models') through research group discussions.

3.3 Results

We observed that patients successfully used self-tracking for managing migraine with a tracking tool that explicitly accounted for goal expressions. This section starts by showcasing key observations (Section 3.3.1). As we unpack these observations, we define terminology (Section 3.3.2), show patients concurrently moved across different stages of tracking for different goals, and share examples of goal evolution (Section 3.3.3). We then present ways patients built upon goal expressions not only to decide what to track in alignment with their goals, but also to recognize when and how to adjust tracking in response to goal evolution (Section 3.3.4). We demonstrate the culmination of these capabilities as highly personalized tracking (Section 3.3.5). We next show patients gained valuable insights and identified where to seek expertise (Section 3.3.6). As a result, they achieved improved understanding and care and were empowered in communication and action (Section 3.3.7).

3.3.1 Goal-Directed Self-Tracking Experience at a Glance

We observed how goal-directed self-tracking supported patients as they configured MigraineTracker for a variety of needs and used it over the course of 383 days on average (min = 269, max = 471). We briefly describe the value this experience brought to patients and demonstrate it through case studies. We then detail themes underlying these case studies in subsequent sections.

All patients found MigraineTracker easy to use, pleasant, customizable, and flexible. They

appreciated its unique features, such as medication targets and retrospective reminders. For example, PT08 said: *"the app's customizability also really helped, because I could add whatever fields I wanted to, and it really felt like the only limit was how burdensome I wanted tracking to be. I definitely made some changes to my tracking based on that throughout the tracking period - it was so helpful that I wasn't locked into tracking any particular field and could alter what I was tracking whenever something occurred to me."* Moreover, patients felt goal-appropriate summaries and visualizations presented the information they needed. They felt empowered in interpreting data, recognizing trends, reflecting on time-bound events, figuring out if medications had an impact, and identifying actions to take. For example, PT03 appreciated the *"crazy charts, reports"*, noting *"It's like I know my head better"*. All patients wanted to continue using the app after the study and preferred the app over their prior tracking experiences. Several wanted to know if the app would be commercially available.

PT02 Aligned Tracking of Different Goals to Her Needs

PT02 started experiencing episodic migraines three years ago. As she was at the early stages of her migraine journey, she set up the app to understand why migraines happened and how to control them. To the former, she included contributors she suspected (e.g., stress, menstruation). To the latter, she recorded if she took rescue medications early enough and how well they worked. She also included items in her tracking routine to reinforce health behaviors that were broadly beneficial (e.g., exercising). PT02 gained insights in relation to some of her goals after collecting data for a few months and reflecting on it. For example, she learned of a relationship between alcohol consumption and her migraines. She continued tracking toward other goals she was still figuring out (e.g., impact of stress) and new goals formed during tracking (e.g., whether a biofeedback device helped).

PT04 Obtained Insights and Adjusted Tracking

PT04 managed migraines along with other chronic conditions such as diabetes. With constraints on side effects and availability of medications, he prioritized learning about contributing activities

and the efficacy of his preventive medication. He recorded presence and severity of headaches along with their context (e.g., levels of stress, amount of sleep, sugar intake). Examining the monthly frequency of his migraines in relation to changes in his preventive medication, he learned the medication did not make much difference. In consultation with his clinician, and considering its negative side effects, he concluded to not continue it. He also learned that stress and lack of sleep were more frequent when his migraines significantly increased in number and severity. He wondered about the potential relation between stress, sleep, and migraines and decided to record stress and sleep on headache-free days in addition to headache days to more fully examine the relation.

PT06 Gained Improved Understanding and Care and Her Goals Evolved

PT06 had no successful prior migraine tracking experience. At the beginning of the study, she recorded her migraines along with various associated symptoms, potential contributors, and treatment information. Consistently tracking for several months, PT06 got better at distinguishing migraines from her everyday chronic headaches as she learned when and how associated symptoms (e.g., light and sound sensitivity) preceded her migraines. Better recognizing migraines led her to take rescue medication sooner, which prevented the migraines from getting worse. Moreover, daily reporting on whether she used different treatment options brought the unexpected benefit of having those options in mind when migraines occurred. PT06 felt her migraines limited her cognitive resources, but greater awareness of treatment options allowed her to apply more when migraines happened. As a result, the average severity of her migraines decreased over the course of the study. Having learned about symptoms and treatments, PT06 was no longer interested in the informational value of tracking them. Nonetheless, she kept the items in her tracking routine as she had other goals. The list of symptoms worked as a checklist for deciding if a daily headache was a migraine and the list of treatment options reminded her of things to do to reduce symptoms.

PT09 Felt Empowered and Sought Clinician's Help

Initially misdiagnosed with cluster headaches, PT09 tracked her symptoms (e.g., their timing and duration) to ensure her clinician had an accurate account of her condition. Tracking information empowered PT09's communication of migraines and helped her feel prepared to discuss care with the clinician. Tracking surfaced areas where PT09 most needed her clinician's input and expertise. For example, tracking data highlighted the high impact of migraines on her ability to function which prompted a conversation about changes in treatments. Moreover, PT09 noticed specific and repeating patterns of monthly frequency and severity of migraines and sought her clinician's input to tease apart different explanations, especially in relation to her preventive medication. Sharing tracking information also led the clinician to learn about PT09's alternative treatments (e.g., marijuana) and to educate her about the potential risks of those treatments (e.g., rebound headaches).

3.3.2 Management, Information, and Tracking Goals

Patients used MigraineTracker for a myriad of reasons, which we organize into management goals, information goals, and tracking goals (Figure 3.6). This categorization was inspired by Schroeder et al. [173]'s categorization of goals, but refines it to capture the range of patient goals we observed in our longitudinal study. We define these goal categories and describe relations among them that shape and drive tracking.

Distinct Classes of Goals. ***Tracking goals*** were goals a specific tracking setup would achieve. For example, 'recording presence or absence of migraines' or 'recording hours of sleep'. ***Information goals*** were knowledge to obtain and questions to answer about one's migraine experiences. For example, 'monthly frequency of migraines' or 'does lack of sleep make migraines more likely?' ***Management goals*** were desired health states to achieve (e.g., 'improved symptoms'), constraints to meet (e.g., 'medication availability'), or needs and values to support (e.g., 'control and agency'). We also observed self-regulating behaviors (e.g., 'holding oneself accountable to exercise') as management goals.

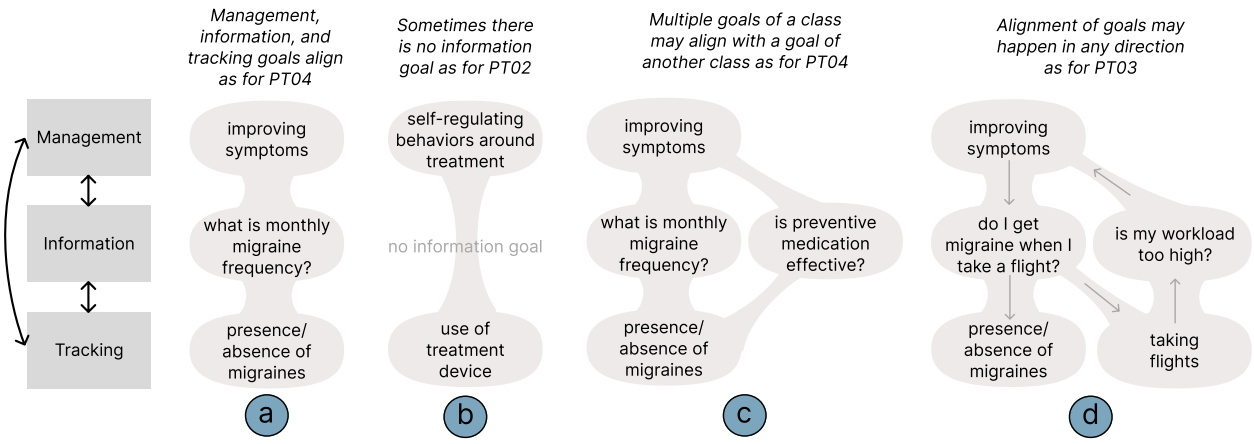


Figure 3.6: Participants described distinct classes of management goals, information goals, and tracking goals as well as different forms of relationship among such goals. An abstract representation of the inter-relation of these goals is illustrated with four specific examples in (a-d). Specifically, tracking goals support management goals (a) with or (b) without information goals. (c) Goals may overlap, such that a goal of one class may relate to multiple goals of another class. (d) There is no strict sequencing in how goals of different classes are related. This characterization of goals is consistent with qualitative and quantitative goals as introduced by Niess and Woźniak [137] and surfaces additional nuance in the inter-relation of goals (e.g., overlapping and dynamic alignments). It also extends Schroeder et al. [173]'s goal types and highlights more complex relations (e.g., goals that concurrently align or evolve).

Tracking Goals Support Information and Management Goals. Tracking goals typically supported an information goal that subsequently supported a management goal. For example, the tracking goal of 'recording presence and absence of migraines' supported the information goal of 'monthly frequency of migraines'. Knowledge of the monthly frequency helped patients such as PT04 make adjustments to treatments or behaviors and eventually achieve the management goal of 'improving symptoms' (described in Section 3.3.1). Another example was using the tracking goals of 'recording migraine severity and duration' to support the information goal of 'how much time is lost to migraine?'. This information goal served patients such as PT09 in their management goal of 'quantifying and communicating' their health state, which might be dismissed because of the invisibility of migraine. All patients used the app to achieve at least one sequence of tracking, information, and management goals. In these cases, data recorded against a tracking goal was of value as it supported the related information and then management goals.

Tracking Goals Support Management Goals without Information Goals. Although tracking goals typically supported management goals through information goals, there were also cases where no information goal was involved. This was most evident when patients had a self-regulation goal, such as exercising, and used a tracking goal to remind and reinforce the relevant behavior (e.g., as with PT02, Section 3.3.1). We observed similar reasoning around other behaviors (e.g., dehydration for PT06 and PT09, stress management for PT02 and PT01). Section 3.3.1's description of PT06 using tracking goals around symptoms to decide if headaches were migraines and her use of tracking goals around treatment options are also examples where tracking goals supported management goals without tracked data enabling an information goal.

Overlapping Inter-Relations among Goals. It was not uncommon for tracking goals to simultaneously support multiple information and management goals. For example, the tracking goal of 'recording the presence or absence of migraine' supported such information goals as 'monthly migraine frequency' and 'how effective are preventive treatments?', which supported the overall management goal of 'improving symptoms'. As another example, PT08 pursued parallel management goals with the same tracking goal. She included a tracking goal for whether she used Cefaly, a neurostimulation device, both to reinforce its regular use (i.e., a self-regulation management goal) and to learn if it improved her symptoms or impacted the efficacy of other treatments (i.e., a management goal of improving symptoms).

Dynamic Inter-Relations among Goals. Classes of goals were not always linked top-down (thus the double arrows in Figure 3.6). There is no hierarchy to imply a strict sequencing of goal types down from management goals, and tracking goals were not always explicitly set toward a specific information or management goal. Similarly, information goals were not necessarily planned according to specific management goals, and tracking goals sometimes inspired new information or management goals. For example, PT02's record of alcohol consumption, tracked as a potential migraine contributor, led her to learn that she consumed more alcohol than expected and highlighted a need for moderation. Neither monitoring alcohol consumption nor its moderation were part of her initial reasons for tracking. Such a pattern where tracking leads

to additional awareness was not uncommon. As another example, PT03 inferred from her records of air travel (i.e., tracked as a potential contributor) that her work had become more demanding, but obtaining information on workload was not initially a goal. The emergence of such new goals is a part of our broader observation of changes in goals within and between classes of goals, as we detail in the next section.

3.3.3 Goal Evolution

Patients had multiple and evolving goals in their use of MigraineTracker, and they could simultaneously be at different stages for each goal. For example, a patient might lapse in one goal while still actively tracking for another, or they might achieve one goal even as they need to continue tracking for others. Although this may seem obvious in hindsight, it is not clearly indicated by current models of personal informatics [111, 63] or goal evolution [137], which tend to consider a single goal for the tracking experience. Designs based on a model of single, separable goals are ill-suited for tracking in chronic conditions in which patients have multiple tracking, information, and management goals, each of which may decrease or increase in priority or resolve at different times, but may share some underlying tracking. In the remainder of this section, we detail participant experiences and some of the complexities associated with multiple evolving goals.

Progression of Individual Goals Across Stages of Tracking. Patients sometimes achieved one goal but needed to continue working toward other goals. This was the case with PT02, who achieved their information goal concerning alcohol but needed to continue investigating the relation between stress and migraines. It was also common to refine a goal or follow up with a new information or management goal. In the latter, patients could abandon tracking the original goal as they started tracking anew for another goal or could concurrently track toward both goals. We observed evolution of information goals for all patients when they reviewed their tracking after several months. Examples include when PT10 refined her information goal from learning about the average length of migraines to learning about the average length of treated migraines (i.e., how

long migraines lasted after taking abortives), when PT06 learned that alcohol was not a strong contributor and moved to investigating 'feeling chilled' as a contributor, or when PT04 followed up on observations of migraine severity by wondering about their daily activities during months with higher severity. New management goals also sometimes emerged after the resolution of information goals. For example, PT01 wanted to prioritize her health over other commitments after learning of her stress-migraine relationship.

Evolution Across Classes of Goals. Changes in tracking goals sometimes followed resolution, refinement, or emergence of information or management goals. Resolution of an information goal usually led to implicit or explicit abandonment of associated tracking goals, especially when those tracking goals did not support any other information or management goals. PT06's removal of alcohol tracking, which was found not to be a contributor, was an explicit change. PT03's lapsing in reporting of brain fog, a symptom she no longer wondered about, was an implicit change. Abandonment did not happen when tracking goals supported new management goals independent of the resolved information goals. For example, PT02 continued tracking her menstruation, even after learning about its connection to migraines, because she wanted to stay aware of that context around migraines near her cycles.

Refinements in information goals often led to changes in tracking goals. For example, PT10 started recording the timing of treated migraines in addition to the total length of migraines. However, there were also cases where no change in tracking goal was needed to support a refined information goal. For example, PT07 wanted a monthly average duration of migraines, after learning about her overall average duration. Although additional processing was needed, she did not need to track differently. Emergence of new information goals sometimes involved new tracking goals, as with PT06's learning about 'feeling chilled'. In other cases, existing tracking goals could adequately support new information goals. For example, PT09's records of migraine presence vs. absence were enough to examine inter-migraine intervals.

Patients sometimes changed their tracking goals without changing the associated information or management goals, often through improvements in their tracking process. For example, PT04

changed his tracking goal of noting the location of pain to recording the presence or absence of pain in frequent locations, a change which simplified his recording. PT07 similarly simplified her recording of sleep from a time range to length of time in hours. The precision of a time range was more burdensome and seemed unnecessary for her information goal of examining the potential link between inadequate sleep and migraines.

Evolution in Goal Priorities. In addition to changes in specific goals, goals became more or less important even as they remained relevant. For example, PT08's migraine severity improved and she then cared more about information goals regarding non-migraine headaches. PT06 cared less about learning if loud sounds contributed to her migraines, as she felt she could do little regarding the source of the loud sounds (i.e., her young dogs) even if there was a relation. Both PT03 and PT09 cared more about goals that implicated a behavior change in day to day life. In addition to actionability, new or unexpected variations made patients more interested in certain information, whereas lack of variation led them to lose interest. For example, PT08 started caring more about migraine duration after noticing increased average duration, whereas PT10 lost interest in the relation of migraines and lightheadedness as it rarely happened.

Changes in what goals were pursued and at what priority often prompted additional changes beyond the classes of goals. The next section details such changes through our observations of tracking and data models.

3.3.4 Expression of Goals Facilitates Alignment of Tracking to Patient Needs

Patients configured MigraineTracker by selecting goals and describing what each tracking item helped them achieve. In doing so, they aligned different management, information, and tracking goals. How patients recorded data during tracking, which we characterize through *tracking models* and *data models*, then complemented alignment of goal types. As goals evolved, goal-centered review additionally facilitated adjustment and re-alignment of goals, tracking models, and data models. We detail these observations by first presenting the tracking and data models patients used and then providing examples of different forms of alignment and

re-alignment. We also point out challenges we observed in the process. Considering the tracking experience in terms of distinct goals, tracking models, and data models, along with the alignment of these elements, offers insights into how goal-directed tracking enabled patients in deciding what to track over time and subsequently highlights support that is relevant to long-term tracking of chronic conditions.

Data Models. We define data models as *units of recording*. Patients used three distinct data models: a ***day-based model*** where recording happens for each calendar day (e.g., presence of migraine or its peak severity for each calendar day), an ***episode model*** where recording happens for each episode of migraine, which may extend beyond a single day (e.g., duration or peak severity for each episode), and an ***interval model*** where recording happens for a window of time, typically since the previous recording (e.g., number of days of migraine within each interval). Although MigraineTracker's default model was day-based, patients adjusted it to other models and sometimes combined multiple data models. For example, PT04 recorded presence or absence of migraines for each *day* but preferred to record migraine duration for each *episode*.

Tracking Models. We characterize *when recording happened* under four types of tracking models: a ***daily model*** where patients recorded every day, an ***event model*** where recording was initiated by an event, often the start of a migraine, a ***divergence model*** where certain changes prompted recording, and a ***hoarding model*** where recording happened occasionally when an opportunity arose (e.g., a break in routine activities every few days). Patients commonly combined these different tracking models. Many patients who reported presence of migraines on a *daily* basis reported associated symptoms (e.g., light and sound sensitivity) in the *event* of a migraine. It was also common for patients to report preventive medication dosage only if they were *diverging* from the previous levels. Recording upon *divergence* from typical levels was also common for stress levels and stressors because of the associated negative affect. PT08 described this: *"I don't want to have to dwell on that, dwell on particular stressors by mentioning them day after day because I don't think that, that would be good for my mental state."*

Aligning Goals, Tracking Models, and Data Models. Patients often used tracking and data models

consistent with their goals. PT01 provides an example of *aligning tracking models to goals*, as she used different tracking models for different goals. She recorded symptoms such as brain fog only on migraine days because she only cared about learning how frequent these symptoms were with migraines. On the other hand, she recorded excessive stress on a daily basis to investigate if she got let-down headaches (i.e., a type of headache that happens when a few days of high stress are followed by release from stress). PT10's recording of alcohol provides an example of *aligning data models to goals*. She wanted to learn if her migraines happened the day after drinking alcohol. Although most of her tracking was interval-based, she used a day-based model for her specific tracking goal of tracking migraines along with prior-day alcohol consumption.

Patients were also able to adjust tracking models as their goals evolved. For example, PT04 switched from event-based recording of stress and sleep to daily tracking in order to more fully understand how stress impacted sleep and how both influenced migraines. Goal priorities also informed tracking models. Some patients had a 'minimum recording set' corresponding to their highest priority goals. This typically included migraine presence, migraine duration, and rescue medication, which patients prioritized recording even when extremely busy.

Mismatches sometimes occurred in aligning goals, tracking models, and data models. The most common mismatch between goals and tracking models was using an event-based model for goals that required daily tracking. For example, PT07 recorded her menstruation only on migraine days, which was inadequate for learning whether and how menstruation affected her migraines. In an example of difficulty aligning certain data models with tracking models, PT01 and PT07 were confused about daily recording of preventive shots they received every one or three months. The influence of each shot extended to other days, so answering 'no' to whether they used that medication felt inaccurate on those other days. The app's default day-based data model also sometimes did not match a patient's data model. For example, PT04 could not directly record the start and end of a multi-day migraine episode. After consulting the research team, he worked around this by adding a custom item 'same as yesterday' to his tracking.

As goals evolved, patients commonly recognized a need for more rigorous tracking models (e.g., from event-based to daily recording). They were less cognizant when goal

evolution meant they could adjust their tracking to be less burdensome. For example, PT06 reached a point where the presence of migraines was the only information she needed to record on a daily basis. Despite acknowledging that she had no need for the data, she wanted to keep recording all information every day and felt doing otherwise was a *"user error"*.

Aligning Classes of Goals. The configuration process for MigraineTracker involved choosing relevant goals and configuring a tracking routine in relation to those goals. In the course of adding and configuring tracking items, patients articulated tracking, information, and management goals and described how they aligned. For example, PT08 included a time of day entry for migraine start time and a text entry for her location when a migraine started (tracking goal). This allowed her to learn when and where migraines were more likely (information goal) to then decide where was safer for her to be (management goal). We similarly observed other patients translating management goals to information or tracking goals and operationalizing the tracking goals into specific settings within the app (see Section 3.3.2 for other examples). Every part of a patient's tracking thus spoke to an explicit need they had.

As patient goals evolved, they repeated the same process to make adjustments to their tracking and re-align it to their needs. For example, after mixed insights on the relationship between sleep and migraines (the information goal), PT06 decided to record sleep in terms of its quality (new tracking goal) and not its length (prior tracking goal). Similarly, PT02 decided to use higher granularity in rating stress levels (tracking goal) after data was inconclusive with respect to the relation between migraine and stress (information goal) at higher levels of stress.

Challenges could arise at various stages of the process from articulating goals, to aligning and realigning them, to retaining the established alignments. Patients sometimes struggled with articulating their goals. For example, PT07 did not initially express interest in knowing which rescue medication worked. As a result, she primarily recorded a list of medications she took and if the overall combination worked for her (which did not allow her to understand the separate impact of the various medications). PT10 did not initially differentiate migraine length before and after taking rescue medication. Encouragingly, this challenge was remediated through goal

refinement and re-alignment of tracking with the refined goals as part of patients reviewing their goals during sessions with the research team.

There were also challenges in alignment of tracking with goals or among distinct goal types. PT10 did not include any items in her tracking routine for the tracking goal of recording persistence of migraines. This meant she lacked data needed to support her information goal regarding efficacy of rescue medication. Review of goals during interview sessions helped surface and address this type of misalignment. In another example, PT03 could not define a tracking goal to support the information goal of determining whether migraines returned because a rescue medication effect wore off or because the medication led to rebound. Because goal expressions highlighted this specific challenge, she was able to seek clinician expertise and analytic support from the research team. Her clinician educated her with general information about the medication, and the research team transformed and restructured her existing data into a goal-appropriate visualization.

There were times when realignment of tracking or goal types to evolved goals could similarly be challenging. With little variation in categorical levels of fatigue (typically 'some' on the scale of 'none'/'some'/'lots'), PT04 could not fully examine the relationship between fatigue and his symptoms. He also did not re-align his tracking through finer-grain recording of fatigue. Existing tracking goals and setup sometimes fell short of supporting new or refined information goals. For example, PT09's new goal was to know times she felt desperate in managing migraines, and she thought the number of rescue medication taken could be a good indicator. Her original tracking, however, could not support this new question because she only recorded whether she took medication and not the dosage. Although goal expression alone was inadequate in both PT04's and PT09's cases, it highlighted what change was needed.

We observed the importance of recording patient rationale for their future use as they articulated and aligned goals, given this may be otherwise lost over long periods of tracking. Information goals were particularly prone to loss. PT08 forgot she chose to record fatigue to understand its variations during migraines. Such loss of an information goal sometimes led patients to lapse in recording. For example, PT07 stopped recording their migraine impact on

disability, which she had originally planned to record to learn the functional severity of her migraines.

3.3.5 Goal-Directed Configuration of Tracking Is Meaningfully Personalized

The ability to express different goals and align tracking to them led patients to highly personalized tracking. Additional examples in this section complement those in earlier subsections to illustrate the importance of the explicit scaffolding of goal expression and the alignment process in facilitating personalization.

Patients customized MigraineTracker's recommended and common tracking items and defined custom items to capture information that mattered most to them. Customization of the app's provided items happened through selecting *how* information was recorded to be more conducive to goals and preferences. For example, patients customized the recording of stress through whether it happened or not (PT01), its qualitative severity (e.g., 'none' vs. 'some' vs. 'lots' for PT02), or a note of what the stressor was (PT08). Patients recorded sleep by its start and end (PT07), as a number of hours (PT06), or as notes of any inconsistency (PT09). Custom items enabled further personalization through defining concepts of interest. Many patients created custom items for specific medications (e.g., ajovy, botox) and alternative treatments (e.g., acupuncture, massage). Custom items were also used to capture person-specific symptoms and contributors. For example, 'clumsiness' mattered to PT08, while PT01 tracked 'numbness'.

Although valuable, customization led to challenges. Patients sometimes forgot how they planned to record information. PT05 forgot he wanted to record his typical rescue medication under 'as-needed medication' and other rescue medication under 'new as-needed medication'. Providing rich support was also particularly challenging for custom items. Patients wanted specific features for the entry and visualization of medication items. For example, PT06 wanted reports of a particular medication to appear with a unique icon on the app calendar view.

3.3.6 Goal Expression Drives Reflection; Reflection Drives Goal Evolution

Reflection happened at the time of recording and when reviewing data summaries and visualizations. In both cases, goal expressions played an integral role. In this section, we first show the connection between goal expression and reflection at the time of recording. Next, we demonstrate that data review around explicit goal expressions enabled reflection. We do so using examples that indicate patients successfully engaged in the processes that Fleck and Fitzpatrick [67]'s and Baumer [11]'s models articulate for reflection. As additional evidence for successful reflection, we show the overall process led patients to gain valuable insights. We then note how the very processes characterized in [67, 11] can also be considered through the lens of goal evolution, and thus advance understanding of the role of reflection in goal evolution [137]. We conclude by presenting challenges patients encountered throughout the reflective process.

Reflection Happened at the Time of Recording. Information goals associated with a tracking goal prompted patients to think about the information they were seeking in the moment of tracking. For example, the act of recording led PT02 to conclude her migraines were related to the amount of alcohol she drank. She had configured her tracking to record whether or not she consumed alcohol, but she did not record alcohol quantity. The insight into relevance of quantity *"was in concert with being really diligent about tracking on the app, but also just having my awareness and my life be very open to what are possible triggers, what are things that are going to potentially lead to a migraine that are within my control?"* Learning from immediate experience while recording also led PT08 to quickly determine that any intense smell could trigger her migraines. For both PT08 and PT02, goal-focused attention at the time of tracking facilitated reflection, which happened either within a short period of tracking or without explicit data entry.

Patients Navigated Levels of Reflection. Patients frequently chose to reflect on information goals that they anticipated would lead to changes in behavior or treatment. As patients reviewed goal-appropriate summaries and visualizations, we commonly observed the first three of Fleck and Fitzpatrick [67]'s levels of reflection: description (R0), description with reflection (R1), and dialogic reflection (R2). Patients commonly attended to patterns that stood out to them (R0),

including minimum or maximum values, consecutive migraine or migraine-free days, and variations over weeks or months. This was often followed by attempts at explaining the patterns (R1), typically in relation to treatments or context including day of week, events, and habits. Patients sometimes considered multiple explanations or tried connecting multiple patterns and explanations (R2). For example, PT04 noticed months with higher frequency and severity of migraines (R0). He then examined the monthly breakdown of contributors to explain the differences and saw higher frequency of stress and inadequate sleep in those months with higher migraine frequency and severity. Connecting the two insights, he next asked if he got worse headaches on days with inadequate sleep, or when he was more stressed, and if inadequate sleep days followed high stress days (R2). PT04's full engagement in the reflective process occurred despite limited experience working with data and even though cognitive load could exhaust him due to his medications.

Breakdown and Inquiry Were Key to Navigating Levels of Reflection. Patient navigation from lower to higher levels of reflection was closely linked to *breakdown* and *inquiry* aspects of reflection [11]. Salient or surprising patterns at R0 signaled a breakdown between patient understanding and data. The inquiry process always ensued to describe and explain the breakdowns, thus patients went to R1 and R2. The process sometimes started with verification. For example, observation of higher likelihood of migraines within three days of taking a medication (the breakdown) led PT03 to first verify how days were counted for 'the number of migraines within three days' of the medication. After the clarifications, she tried explaining the pattern and considered multiple explanations: *"whether it's a rebound headache and that [the medication] caused the headache or if it just wore off and those headaches days are still there."* Hypothesis formation was integral to the inquiry process, and sometimes relied on defining new concepts. For example, PT09 noticed patterns of migraine and migraine-free days and wondered if there was a fixed 'inter-migraine interval': the number of consecutive migraine-free days between consecutive migraine days. The process of noticing breakdowns, defining concepts, and forming hypotheses led patients to form complex information goals they had not previously considered.

Reflection Brought Value. Goal-appropriate summaries and visualizations addressed some of the most pressing information goals that were initially set or arose from the inquiry process. All patients found insights that addressed needs (e.g., around symptom patterns during the week; medication efficacy; or key contributors such as sleep, stress, or alcohol). Moreover, patients identified when they needed additional data or expertise. For example, PT04 learned he needed to record sleep and stress on both headache and headache-free days to fully establish a hypothesized relation that increased stress caused decreased sleep which led to migraines. PT05 decided to integrate diet tracking data to cross-check the specifics of his diet and headache patterns. Insights from reflection also guided patients in seeking expertise, particularly from their clinician. PT08 wanted her clinician's advice for dealing with multi-day migraine episodes she learned were common for her. PT03 wanted her clinician's input on alternative hypotheses about a medication (if it caused rebound headaches vs. its effectiveness ran out). Patients also sometimes looked to online resources to follow up on or augment insights from reflection. PT09 wanted to know how typical her migraine duration and frequency were.

Patient goals evolved as they went through levels of reflection, defined concepts, formed hypotheses, and gained insights. Defining concepts and forming hypotheses led to new or refined information goals, as with PT03, PT04, and PT09's above. Gaining insights addressed existing information goals and was sometimes followed by new goals, as with PT05's above. Patients then aligned tracking goals, tracking models, and data models with evolved information goals (Section 3.3.4). Resolution of information goals was sometimes followed by new management goals, as in PT06's use of light and sound sensitivity to make sense of her everyday headaches and whether or not they were migraines.

Challenges. There were pain points in the reflection process. Patients sometimes struggled to form hypotheses. PT04 was unable to explain the weekday differences in migraine frequency. Clinicians were often a good resource. Brainstorming with their clinician led PT04 to consider increased social interactions during weekends as a potential explanation. Closely related to the challenge of forming hypotheses was the challenge of developing hypotheses. MigraineTracker's

emphasis on expressing goals and aligning tracking with goals meant developing hypotheses, where goals were yet to be well-defined, did not receive much support. As PT06 learned about week of month differences in migraine frequency, she said *"food for thought, is what this is. This now makes me want to start paying attention to what else is going on, …I don't know yet what I would want to record to go along with this, but it just makes me curious, and I'll probably have to sit and think about it, and maybe even take a month or two to observe on my own before I go, "Okay, here's something I want to start watching," and then stick it into the app."* Patients sometimes forgot insights they gained from their reflection. For example, PT08 paused a preventive medication mid-way in her tracking, learned of increased migraine frequency, and resumed the medication. However, she could not explain the sudden increase then decrease of migraine frequency when she considered her data at the end of tracking. Another challenge was in identifying complex patterns. For example, PT08's data showed higher everyday headache severity followed within a few days of higher stress, but neither PT08 nor her clinician were able to identify this relationship with the available visualizations. Patients also misinterpreted data. PT10 interpreted higher likelihood of migraines when not drinking alcohol to indicate that alcohol helped. She was surprised but did not consider alternative explanations or additional factors (e.g., that alcohol consumption typically happened on good days and in the absence of other contributors).

3.3.7 Goal-Centered Insights Enable Understanding, Communication, and Action

Goal-directed tracking led to practical insights, informed behavior, and facilitated help-seeking and communication. We observed these benefits as well as challenges and additional considerations in effectively supporting them.

Insights from the tracking experience led patients to better understand and manage their condition. For example, commenting on improved migraine severity over the course of the study, PT06 said *"I think I'm getting better at nipping them in the bud. The app has helped me… It's helping me recognize when I have one sooner, and helping me just go ahead and take the damn drugs. So they're not getting up to six, seven, eights and nines… I am actually glad to see that. I'm glad to see that I'm taking better care of myself. I'm not suffering."* Patients identified whether

changes they had made helped or if they needed to make further adjustments in daily behavior or medications. For example, upon seeing improved migraine frequency, PT01 decided to continue her new preventive medication. Comparing the efficacy of different rescue treatments convinced PT09 to take naproxen less and rizatriptan more. Patients sometimes identified needs they had not otherwise considered, as with PT01's realization of a need for prioritizing stress management.

Added understanding of their condition facilitated improved patient communication with clinicians. For example, PT05 felt the information helped *"accurately express things that I've wanted to express to a doctor"*. PT09 similarly found that reviewing goal-appropriate material helped her *"feel more prepared to see my provider"*. Clinicians also described how tracking and a focus on goals prepared patients to take the lead in conversations, making more effective use of their sessions. PR02 described that ideally *"we'll have a conversation, and then that conversation will lead to, "What can we do about it?""*. This ideal vision was realized in her interactions with patients through goal-directed tracking. Commenting on the session with PT08, PR02 noted *"having her tracking, she already had it in her head, what might be contributing? And so then, we could have this full on conversation about, "Okay, how do we change this?""*. In leading conversations with clinicians, patients sought expertise where they needed it the most (e.g., in the challenging task of translating insights to actions). For example, PT04 was able to get advice on reducing migraines on weekends by taking breaks from social interactions. PT08 worked with her clinician toward a concrete plan for ensuring adequate sleep (i.e., having dinner earlier).

PT09 also saw opportunities to use data and analysis from her tracking to communicate with others: *"I just wish people could see this in my academic and professional life, and also in my personal life. You have to disappoint a lot of people when you have migraine by canceling, not being available, calling out… I do wish that I could show them, I can't be there and I failed to be there because of this, what we're seeing here."* She felt the insights helped her better advocate for herself.

Not all insights led to action. Patients were more likely to act if they felt the underlying relationships were strong enough relative to the cost and feasibility of taking action. For example, the relationship between migraines and loud sound in PT06's view *"is not huge. 26% is not a big number. And with the house that I live in… It's just noisy."* (Figure 3.4, c). Perceived necessity of

action also influenced how patients reacted to insights. An example is PT08 who did not feel a need for further changing treatments because of already-achieved improvements in her condition. Backed by PT08's data, which still showed high migraine frequency, her clinician was able to talk with her about additional change.

Despite gains in communication and action from goal expressions and subsequent insights, patients needed further support. Although they were empowered to seek clinician expertise for action planning, they were mostly on their own in following through with actions. For example, both PT01 and PT02 included their use of a biofeedback app recommended by the clinician as part of their tracking routine. Despite sincere intentions, neither followed through with actually using that app, as being reminded of it while tracking was not sufficient.

3.4 Discussion

We studied the lived experience of goal-directed self-tracking. Patients described distinct and evolving goals for self-tracking related to their migraines and concurrently pursued those goals across distinct stages of tracking. Our observations extend prior work on goals and goal evolution [137, 173] in detail and scope. Goal-based support in MigraineTracker and accompanying visualizations facilitated awareness of and progress toward qualitative goals. Moreover, we concretely illustrated results past work speculated and anticipated (e.g., goal evolution) and uncovered how these are achieved (e.g., reflection and realignment). Specifically, we observed the importance of scaffolding around expression of patient goals to ensure goals were aligned to each other and to other aspects of tracking, including tracking models and data models. This led to a highly personalized tracking experience. Goal expressions also facilitated reflection, which improved understanding, communication, and action. Reflection in turn drove goal evolution. Overall, expression of goals enabled patients to externalize their needs and values and situate tracking accordingly. As predicted by prior work [15, 167], this led to improved sense-making and condition management. Below, we discuss how identifying distinct goals provides an analytical lens for analyzing and designing personal informatics systems (Section 3.4.1). We also note the importance of goal-specific adaptations of existing

models of personal informatics (Section 3.4.2). We then share design implications of our observations (Section 3.4.3) and reflect on our methods and their limitations (Section 3.4.4).

3.4.1 Classes of Goals as Analytical Tools

We identified distinct classes of goals in patient use of MigraineTracker. This distinction between management goals, information goals, and tracking goals provides a novel perspective for understanding self-tracking in chronic condition management and for designing effective support. Not accounting for these goal types and their interconnections leads to design gaps. For example, the need for aligning different goal types cannot be recognized without first distinguishing goal types. Goals such as self-regulation are unlikely to receive adequate support if we overlook tracking goals that may exist without an information goal. Our observations suggest no goal-directed tracking tool can be expressive enough unless it supports an interconnected subset of management goals, information goals, and tracking goals. Related to and in consideration of the range of goals we observed in each class, it is reasonable to expect that any design may be incomplete in what goals it anticipates. It is highly likely to encounter unknown management goals, advanced information goals that rely on unsupported analysis, or unconventional tracking goals for recording information in new ways. Considering distinct classes of goals can guide development of specialized designs that should be in place for a successful tracking experience. It can also inform the flexibility we should aim for in designs to enable people to adapt their tools to their evolving goals.

Considering distinct classes of goals also provides an analytical lens to understand the failures and shortcomings of existing tools. For example, a design that only supports an event-based tracking model will fall short in supporting people in achieving goals that rely on a daily tracking model. People may still be able to appropriate tools if a design does not undermine their ability to do so. For example, a tool with an event-based tracking model that allows certain entries to be left blank might be appropriated for daily recording.

We emphasize that we are not the first to note different goal types. We complement prior work, such as [137, 173], by bringing new and more detailed understanding of goal types and how

they relate to other aspects of tracking. By elaborating upon goal distinctions and inter-relations we draw attention to areas to which designers and researchers should attend.

3.4.2 Accounting for Multiple Goals in Models of Personal Informatics

Existing models of personal informatics [111, 63] and goal evolution [137] describe the self-tracking experience with the unstated assumption that stages and concepts apply to a person's entire tracking experience. Existing models do not strongly distinguish among the various interrelated goals a person may be pursuing through tracking nor depict how those goals may change or resolve at different times. Our observations demonstrate that people are simultaneously at different parts of these models for different goals. For example, because a tracking goal may support multiple information goals, confusion can arise when an information goal resolves or changes but other information goals continue, or similarly when a patient lapses in one information goal but continues with others. It can be unclear what the person needs to continue tracking or what might they stop tracking or change to tracking using less burdensome models.

Considering goal-specific versions of existing models of tracking can inform future system design and analysis. For example, consider the preparation stage of tracking, where people decide what and how to track. A goal-specific design could support different tracking models for each goal rather than assuming a fixed model for all goals. Another example is lapsing. A goal-specific design could better support different forms of lapsing [173] with goal-specific support (e.g., goal-specific vacation-lapsing, informed in part by limitations of MigraineTracker's support for vacation-lapsing across all tracking). MigraineTracker's design for intentional lapsing in tracking was based on insights from prior models of personal informatics [63]), but was ultimately inconsistent with patient experience (e.g., there was a minimum set of tracking goals that patients maintained even during vacation). Considering models of personal informatics at the level of goals also highlights the need for accounting for how evolution of one goal is related to other goals. For example, reflecting on one goal may influence the preparation, collection, and interpretation of another goal. Tools that account for

such inter-relations and facilitate people in adapting their tracking as goals resolve or evolve can provide a better tracking experience.

3.4.3 Implications for Designing Goal-Directed Tracking Tools

Explicit scaffolding around expressing distinct goals and aligning them to each other played a key role in personalization of tracking and gaining value from it. Prior work had highlighted the importance of explicitly supporting the initial articulation of goals [173]. Our observations of goals over an extended period of tracking underscore the importance of open design problems in supporting not only an initial articulation of goals but also their alignment and re-alignment. Our prototype supported initial goal articulation by offering a list of options from which people could select. However, there were cases where patients did not accurately articulate goals (e.g., PT07's and PT10's struggles in initial articulation; Section 3.3.4). Well-designed goal-setting practices involve feedback loops and opportunities to adjust goals [118], and our results add to literature calling for the HCI community to develop design practices that support reflecting on and revising goals over time [3, 56] (e.g., using techniques for scaffolding the process and targeting opportune times for reflection so as to avoid rumination, a potential unintended consequence of self-tracking in which people dwell on negatives and blame themselves rather than finding potential solutions and experiencing progress [55]). As noted in prior work [37] and as we observed in our study, reviewing and updating goals before a clinical encounter is one such opportune time. It can focus the visit and make the most of a clinician's expertise (e.g., rather than retreading things the patient already knows). Updating goals during a clinical visit can also incorporate clinician expertise into goals going forward. We further note that as people gain experience with tracking and their priorities change, they may also wish to revisit and tune their goals and associated tracking routines, a process that designs should explicitly prompt and support. Additionally, goal expression alone was not always sufficient for aligning goals, tracking models, and data models (e.g., PT03's, PT04's, and PT10's challenges in goal alignment; Section 3.3.4). Reviewing goals, especially after some tracking data had been collected, and with clinical expertise or tracking and analytics expertise (often provided by the

research team in this study), could help detect and correct misalignments. As our study relied on human resources that may not be available to everyone engaging in tracking (clinicians and researchers), future work should develop design strategies for supporting this review process. This might include structured walkthroughs (e.g., through conversational agents [107]) or review interfaces (e.g., dashboards or visualizations that can highlight misalignments between expressed goals and data being tracked).

Our results also highlight opportunities for specialized support of different management, information, or tracking goals related to similar data or activities. For example, consider recording exercise to self-regulate vs. to learn of its relation to migraines. In the rare event that similar tracking and data models apply to both cases, other aspects of their support could be different: goal realization techniques such as implementation intentions [74] would be more appropriate to integrate in reminders for a self-regulation goal and less so for a learning goal. Visualizations for a self-regulation vs. learning goal could also vary in complexity. Enabling specialized support depends not only on eliciting goal expression but also accounting for nuanced inter-relations and evolution. Tracking tools can adapt support to such specifics of the goals and can capture this information instead of relying on an individual's memory, where we saw it was prone to loss over time.

Reflection is integral to goal evolution. Goal-directed self-tracking tools should therefore address difficulties that impede the reflection process, including forming or developing hypotheses, identifying complex patterns, and misinterpreting information. Designs can leverage clinician expertise, new interactions, and computational techniques to better support these tasks. For example, tools could enable brainstorming with clinicians for hypothesis formation or could use mixed-initiative pattern discovery techniques to surface complex patterns that might otherwise be missed, similar to [37, 171].

3.4.4 Reflections on Methods and Limitations

We used MigraineTracker as a technology probe to understand patient experience with goal-directed tracking in the wild. The benefits that it brought to patients should however be

considered in the full context of our study. For example, we asked patients to think aloud as they configured MigraineTracker and to explain why they made various selections. This aspect of our method led them to describe different goals and ensure they had a setup consistent with those goals. We also had patients repeatedly review and comment on their goals as well as whether and how their data supported those goals. Although primarily intended to elicit feedback, these aspects of our method substantially influenced patient experience and suggest opportunities for future designs.

We intentionally offered data review via simple static visualizations. Consistent with Moore et al. [133]'s insights, this approach is both more cost-effective than building a custom exploratory data analysis tool in a poorly understood design space and more conducive to generating truthful design requirements. Keeping the data presentation simple and static focused our sessions with patients on *what they want* to achieve, instead of being distracted by comments on the usability of visual elements.

We analyzed our data primarily from a patient perspective, to center their goals and ways in which MigraineTracker did or did not support them. Future analysis should more deeply examine the patient-clinician interactions in our study as well as clinician experiences to identify ways in which the design supported collaborations and clinician needs and to surface additional opportunities for better support. Although models of personal informatics that were centered on the individual [111, 63, 137] facilitated a patient-centric and goal-centric analysis, examination of the collaboration might also draw upon other lenses, such as patient-generated data as a boundary negotiating artifact [36]. As prior research has emphasized that tracking to manage a chronic illness is a process with many interested parties (e.g., clinicians, family members, informal carers, workplace and community members) [190, 135, 151], future research might also examine the ways that goal-directed self-tracking technologies can support communication and coordination across a broader range of parties.

3.5 Contributions to Thesis

This chapter demonstrated goal-directed design and scaffolding around explicit expression of goals offers a pathway to addressing challenges in deciding what and how to track, how to adjust tracking, or how to interpret data. We observed expression of goals facilitated externalization of distinct goal types and alignment of these goals to each other and to the specifics of when and how recording occurred. Patient tracking was highly personalized to their needs as a result. Goal expressions also supported reflection through goal-appropriate material, and reflection in turn led to goal evolution and enabled improved understanding, communication, and action. We highlighted the importance of accounting for distinct goal types in the design and analysis of self-tracking tools and highlighted the need for goal-specific adaptations of personal informatics models. Our findings also underscored the requirements for effective goal-directed design and the need for further research addressing these requirements. We need to support alignment, evolution, and reflection in goal-based systems. In the next two chapters I introduce ideas exploring this design space.

Chapter 4

SELF-REGULATION OF GOAL PURSUIT FOR WORK-NONWORK BALANCE GOALS

Findings of Chapter 3 highlighted key activities in goal-directed self-tracking. These include expressing goals, examining and refining goals as they evolve, aligning and realigning goals with each other and with tracking and data models, as well as planning goal-aligned actions. However, carrying out these activities is not trivial. In case of action planning, many people struggle with starting to act, staying on track, and adjusting their efforts and resources [75].

Consistent with the broader literature [111, 63], findings of Chapter 3 also underscored the importance of reflection on personal data in goal-related activities including action planning. Reviewing and interacting with data allows people to navigate stages of reflection from revisiting their behavior patterns to explaining them to exploring relationships [31]. This process increases people's awareness of their behavior and leads to insights [31] that are instrumental in making decisions on the actions to take [111] and in coming up with personalized plans [109]. However, reflection is itself challenging and support for it remains inadequate [29]. Although explicit expression of goals guided reflection, additional scaffolding is needed.

This chapter examines one such scaffolding by drawing on implementation intention (II) and mental contrasting (MC) techniques, two self-regulation techniques of goal pursuit that have been shown to help individuals in action planning [138]. IIs are 'if-then' plans connecting a critical situation to goal-directed actions that help one achieve the desired goal (e.g., "if I crave sugary snacks, then I will eat a healthy fruit instead" for the goal of "healthy eating"). MC asks people to elaborate on their desired state and identify the obstacles standing in the way of realizing that state (e.g., "disorganized notes" is an obstacle to the desired future of "excelling in exam"). II and MC have been successfully used in a variety of settings and are especially

powerful when used in combination [138], where if-then plans are formed around obstacles (if-part) and actions to prevent or overcome them (then-part). However, whether and how these techniques can support reflection on personal data remains unexplored. Past HCI research has primarily considered enhancing the *execution* of II plans via reminders [152, 16] or recommending automatically generated II plans based on personal data [51]. *Forming* the behavior plans as a part of a reflective process has not been considered.

This chapter therefore examines the use of IIMC to scaffold reflection on personal data for action planning toward improving work-nonwork balance, where the complexity of the context and demands intensifies the challenges of action planning [141]. Specifically, it considers the following research questions:

RQ1 Does IIMC-based reflection on personal data improve behavior planning via enhanced understanding of obstacles on the way of work-nonwork balance and the opportunities to respond to these obstacles?

RQ2 Does IIMC-based reflection on personal data improve perceived work-nonwork balance via enhanced behavior plans during the study? If so, does the improvement surpass that of alternatives?

To address these research questions, I designed a system that allowed people to collect personal data and reflect on it as they created behavior plans based on IIMC instructions. I then evaluated this system in a three-week between-participant study with 43 participants in four groups: the group who used data-driven IIMC planning (D+IIMC) compared to groups who used only data (D-only), only IIMC instructions (IIMC), or neither (basic control). Findings show that D+IIMC group significantly outperforms others in several tasks pertaining to the pursuit of work-nonwork balance: with D+IIMC, people are more aware of their plan compliance and more capable of rescheduling their plans as personal or socially-imposed changes arise. Findings additionally highlight the importance of reflection on the recent past (micro-reflection) in improving these tasks as well as measures of work-nonwork balance and time-management.

I discuss implications of these findings in terms of future design opportunities to support IIMC-based reflection on personal data in work-nonwork balance context, along with considerations for the collaborative and multi-stakeholder nature of this context. In summary, the contributions are:

- I propose *IIMC-based data review* to support goal-aligned action planning. I design WoNoB[1], a tool that realizes this technique in drawing insights from personal data that guide individuals in deciding what actions to take as well as when and where to take them toward their goals (Section 4.1).
- I evaluate IIMC-based data review through a three-week between-participant experimental study with 43 participants (Section 4.2) and confirm that IIMC scaffolding for reflection improves performance on such tasks as plan compliance awareness and rescheduling (Section 4.3).
- I present design implications and considerations for combining IIMC and data review (Section 4.4).

I completed this work along with Matthew Jörke, Jina Suh, Koustuv Saha, Shamsi Iqbal, Gonzalo Ramos, and Mary Czerwinski, where I led the design of WoNoB system, study and its material, preparation, planning, and coordination of the study, study sessions, data analysis, and manuscript preparation in close with the team. I will subsequently use first person plural to describe the work and its contributions. This work was published in ACM Proceedings on Human-Computer Interactions (CSCW'24) .

4.1 System for Data-Driven IIMC

Our objective is to examine if and how the use of implementation intention and mental contrasting (IIMC) to scaffold reflection supports decision-making in realizing work-nonwork balance goals, whereby people draw insights from their personal data to decide *what* actions to take as well as *when* and *where* to take them. We do this by developing WoNoB, a system that

[1]Read *'wanna be'*, an acronym for work-nonwork balance

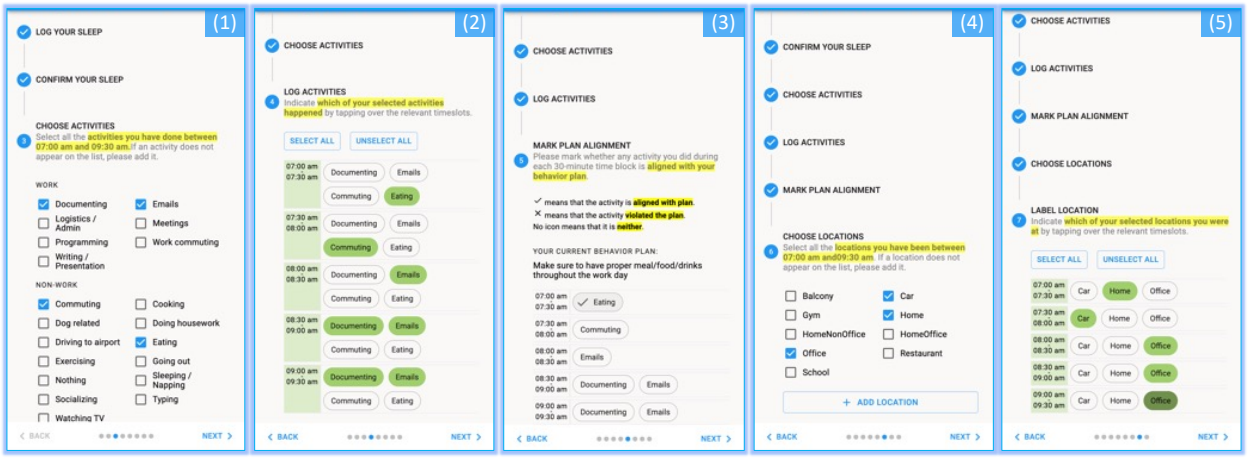


Figure 4.1: Interface for active reporting of activities, locations, and whereabouts. From left to right: (1) participants selected the activities they engaged in within the specific time window (7:00am to 9:30am in the example) from a customized list of activities. (2) They then marked the activities that happened during each 30-minute time slot (second to left) and (3) decided if their activities were aligned with their behavior plan ('Make sure to have proper meals/food/drink throughout the work day' in the example; middle image). (4) Next, they selected relevant locations to (5) mark per 30-minute timeslots (last two images).

allows individuals to review their personal data to examine their work-nonwork balance state, identify obstacles that hinder their desired state, and discover opportunities to prevent and overcome obstacles. WoNoB consists of (1) an active reporting tool (Section 4.1.1) to collect data on activities, whereabouts, and progress toward work-nonwork balance plans and (2) a behavior planning tool (Section 4.1.2) that facilitates reflecting on data within the IIMC framework.

4.1.1 Active Reporting of Activities, Locations, and Progress

We designed a mobile-friendly web interface for active reporting of daily activities and whereabouts as well as progress toward plans for improving work-nonwork balance (Figure 4.1). Participants customized the interface with the common activities and locations that were applicable to them on a day-to-day basis. They could also update these settings whenever a new activity or location became relevant. A Microsoft Teams chatbot reminded participants five times a day during waking hours to log information over 30-minute intervals for the past 3-4

hours. We allowed back-filling of information for up to a day in the past if participants missed the chance to react to the reminders. Past work on experience sampling informed our choice for the frequency of reminders and the window of logging. Specifically, five daily experience sampling reminders enable high recall and are not too frequent to be disruptive over extended periods (three weeks and above) [34]. Moreover, recalled activities within a day are a reliable representation of time use [92]. The choice of 30-minute granularity kept reporting burden less than 2 minutes per 3-4 hours of reporting, the typical interval between consecutive reminders.

We collected activity, location, and progress data to enable IIMC-based reasoning where individuals need to understand their actions and the context for those actions. We chose active reporting over passive sensing of data (e.g., through automatic activity or location detection) as active data better support us in establishing whether or not IIMC-based reflection on personal data can help action planning. Active reporting minimizes the risk of inaccuracies in passive sensing or data errors, which would compromise our ability to conclude if scaffolding reflection with IIMC is beneficial. Active reporting is a time-intensive activity that is tenable only for short periods of time, usually in experimental settings [188]. However, given the exploratory stage of our research, we aimed to investigate the feasibility of our approach under ideal circumstances and collect insights on requirements to strive for under more realistic situations (e.g., data granularity, or necessary data sources). In Section 4.4.3, we discuss the potential for our approach to be augmented with passive sensing.

4.1.2 Behavior Planning through Reflection on Historical Personal Data

We designed our desktop-friendly planning tool to help individuals understand their work-nonwork balance state and identify opportunities and actions to improve it. It comprises four key elements: instructions, filters, calendar, and summary (Figure 4.2):

Instructions. We provide scaffolding for reflective thinking as a series of instructions that appear on the left side of the interface (Figure 4.2a). WoNoB first invites people to think about their desired work-nonwork balance state. It next asks them to use their data to consider their current work-nonwork balance state against the desired state, identify obstacles in the way of

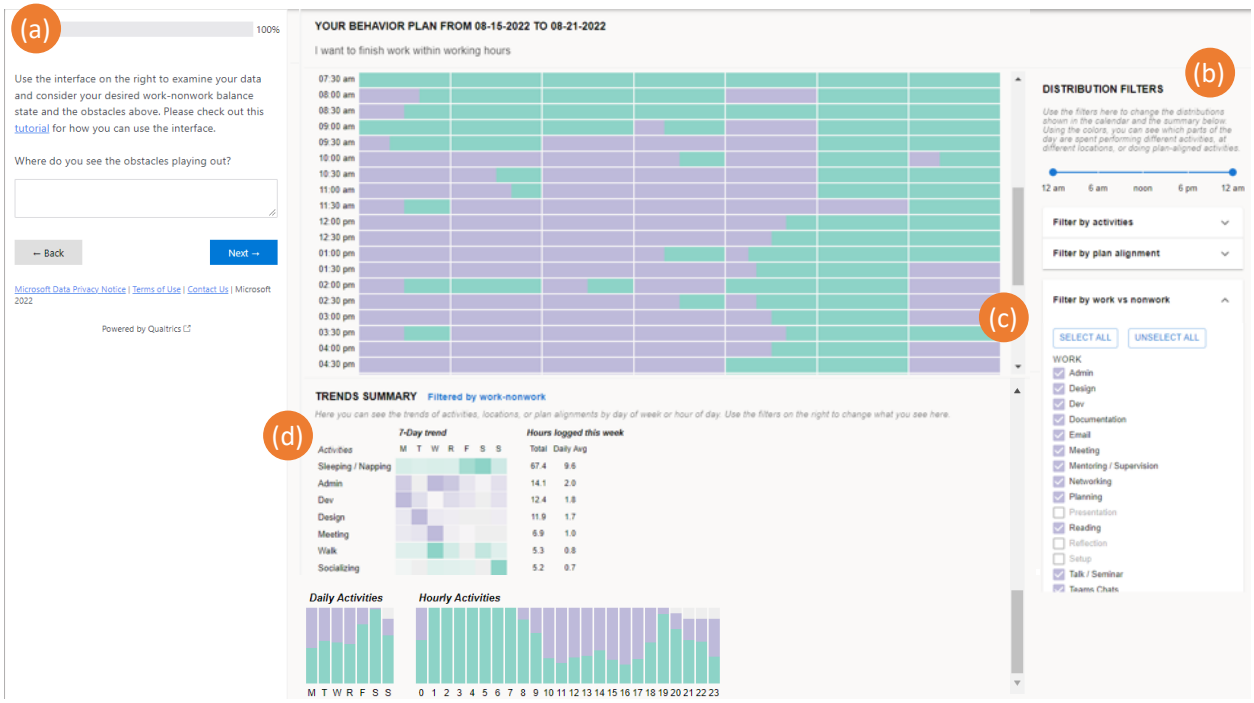


Figure 4.2: Data-driven behavior planning interface. (a) IIMC instructions were given on the left. (b) Participants could choose one of work vs. nonwork, activity, location, or plan alignment filters to get different views of their data. (c) The calendar in the middle displayed data of interest across days of the week and times of the day. In work vs. nonwork view, purple slots represent work while green slots represent nonwork. Slots are split into purple and green halves if activities of both types were reported in them. (d) The interface displays different summary information below the calendar. These include total and average reported hours, distribution of time spent on activities each day (the longer the time, the darker the day under '7-Day trend'), across the week (the 'Daily Activities'), and across the day (the 'Hourly Activities'). Observe that both calendar and Hourly activities show work hours typically start between 8-9 am on workdays.

achieving their desired state, and explore opportunities for addressing these obstacles. It then guides individuals to use these insights for creating if-then statements that describe behaviors for preventing or overcoming obstacles. This process is detailed in Figure 4.3.

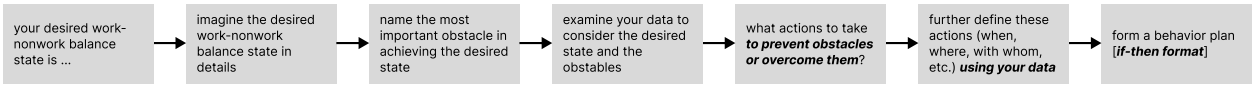


Figure 4.3: Instructions for data review using implementation intention with mental contrasting.

Filters. We enable participants to select and view their data based on aspects such as times of day, specific activities, work vs. nonwork, locations, and progress along their plans through various filters on the right side of the interface (Figure 4.2b). Calendar and summary information is updated per the choice of filters (see Figure 4.6 and Figure 4.7).

Calendar. The interface displays the distribution of the filtered data across days of the week and times of the day (Figure 4.2c). Slots are color-coded based on the content being viewed. For example, work slots appear in purple while nonwork slots appear in green under work vs. nonwork filter.

Summary. Aggregate information over the week (total and average time), across days of week (7-day trend and daily graph), and hours of day (hourly graph) appear below the calendar (Figure 4.2d).

Design Rationale

We incorporated the above elements following different design patterns that previous work had identified as beneficial for reflection [14]. These include visualizations (calendar and summary), statistics (summary), textual prompts and questions (instructions), and refining and revising different aspects of data (filters). Our design assumes the availability of personal data about an individual's activities, whereabouts, and progress, operationalized as alignment with and violations of existing plans for achieving goals (see Section 4.1.1 for details on the interface we used to get this information for the purposes of our study).

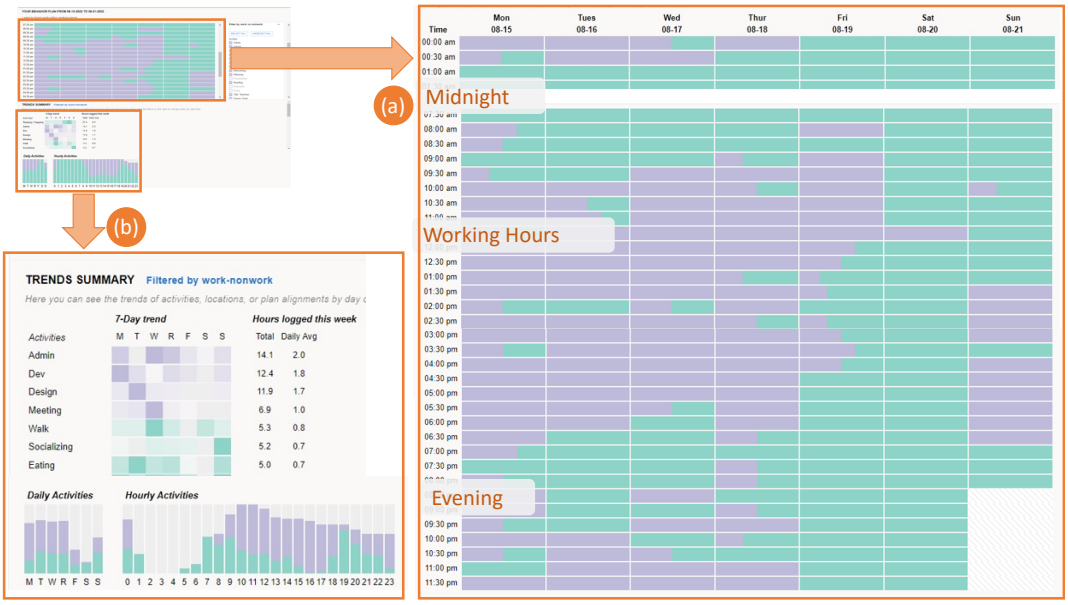


Figure 4.4: Work-nonwork balance view of how time is spent. (a) Work (purple slots) appear after midnight, during working hours, and into the evening. (b) Admin then Dev are the top activities in total hours logged (14.1 and 12.4 respectively).

Walkthrough

Below, we demonstrate the use of different elements of the planning tool with a walk-through. For example, let us follow a software developer whose goal for improving work-nonwork balance is to ensure they finish work within working hours and spend more time on non-work activities, such as listening to the audiobook of a fantasy novel. The instructions on the left (Figure 4.2) ask the individual to draw from their data to (1) consider their desired work-nonwork balance state as well as their current behaviors, (2) identify obstacles to finishing work in time and carrying out nonwork activities of interest, then (3) come up with a behavior plan to prevent or overcome those obstacles, i.e., to decide *what* actions to take as well as *when*, *where*, and *how* to take them. The individua is specifically instructed to examine how and where they spend time and how well their behaviors align with their plans.

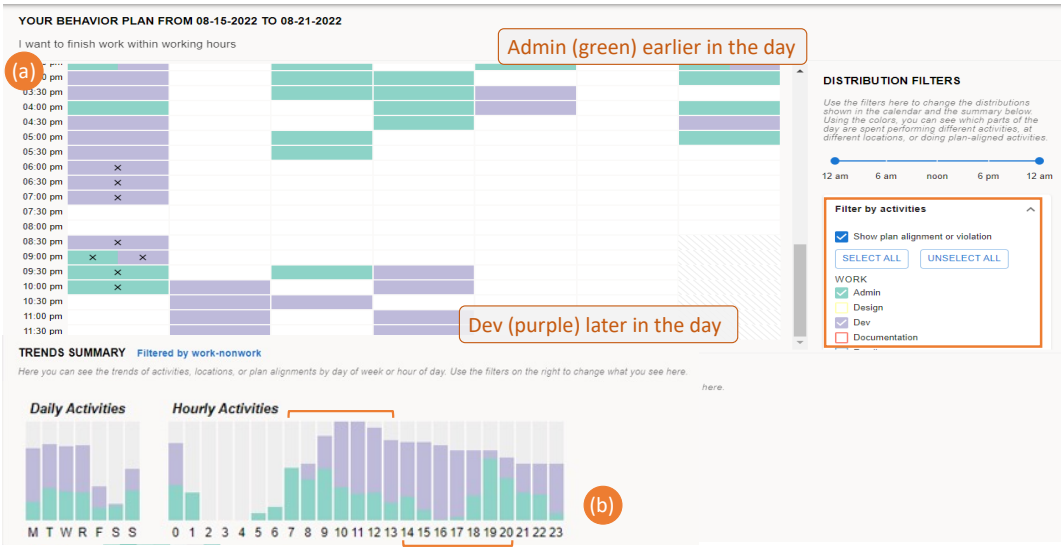


Figure 4.5: Activity view of how time is spent. Only 'Admin' and 'Dev' are selected of all activities in the activity view filter. (a) Admin (green slots) clusters earlier in the day, while Dev (purple slots) appears later in the day. (b) Larger green segments before 10am indicate more time is spent on Admin than Dev early in the day. Dev work is most prominent in later hours of the afternoon (6-8pm).

How Is Time Spent? There are multiple ways available to the individual to examine how time is spent. Filtering data by work vs. nonwork, we can see work is happening after midnight, during work hours, and into the evenings in our example (Figure 4.4a). 7-Day trends offer a ranked summary of time on various activities. Administrative (Admin), then development (Dev) work takes most of the time for the hypothetical software developer (Figure 4.4b). Filtering by these two activities, one can see administrative work slots clustering earlier in the workdays and development work slots usually starting later in the afternoon (Figure 4.5a). The hourly activities chart corroborates this observation, where larger segments for administrative work appear during early work hours, whereas larger segments of development work appear during late evening hours (Figure 4.5b). These observations highlight an important obstacle to finishing work on time: spending too much time too early on less important work leads to working in the evenings or even after midnight.

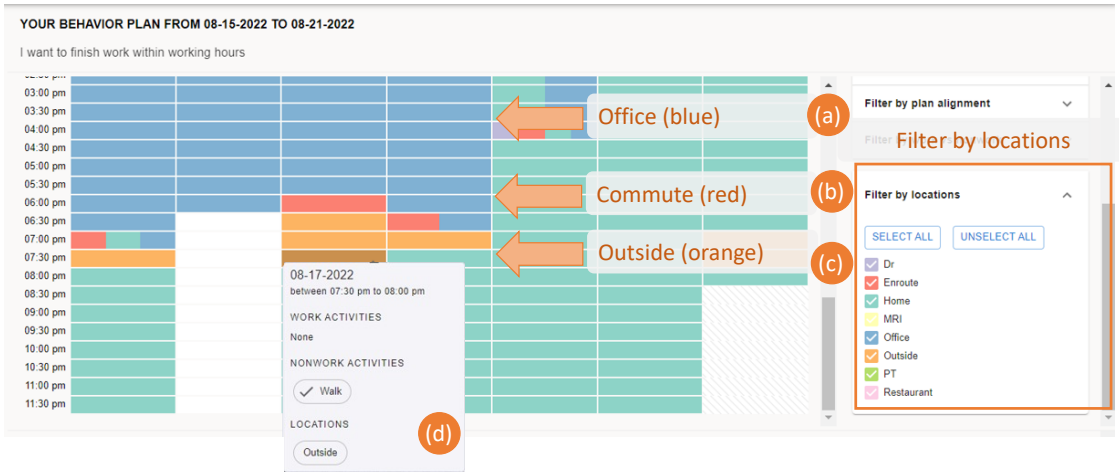


Figure 4.6: Location view of where time is spent. (a) Blue slots show time spent at office during working hours. (b) Red slots indicate up to 30 minutes of commute 6-7pm. (c) Orange slots suggest time is spent outside after the commute. (d) Walking is the reported activity for time spent outside when hovering over the orange slots.

Where Is Time Spent? Filtering data by location, we can see typical places one spends time at different hours (e.g., being at the office during work hours on workdays; Figure 4.6a), timing and duration of commutes (Figure 4.6b), and patterns of transitions from one place to another (e.g., spending time outside after office, Figure 4.6c). Hovering over the time slots to examine the activities we can additionally see common activities at a location (Figure 4.6d). In our example, walking is the most common activity when spending time outdoors. Knowledge of whereabouts and common activities can help carve out time for planning activities that are not currently happening. The software developer in our example might decide to listen to audiobooks while walking.

Do the Existing Behaviors Match the Desired Behaviors? Filtering data by plan alignment, we can examine the timing and distribution (Figure 4.7a) of both alignments and violations of the existing plans for achieving work-nonwork goals. We can further examine aligning/violating activities by overlaying this information within filter by activities (Figure 4.7b). In the example, there are no violating behaviors when the end of the work day

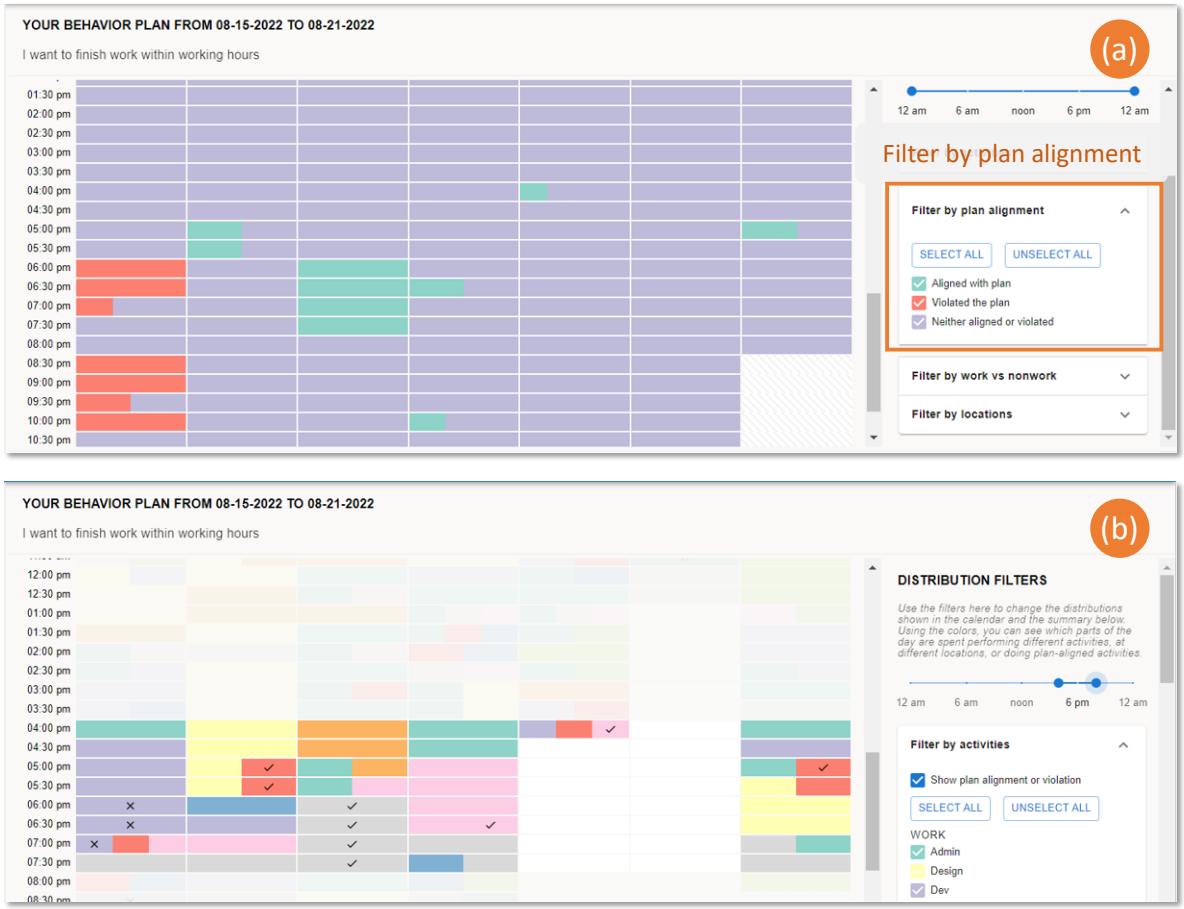


Figure 4.7: View of alignments and violations of planned behaviors. (a) Alignment is reported for several days after 5pm under plan alignment view. (b) Examining the window of time where alignments are reported under activity view, shows a recurring pattern of gray slots (documentation).

is spent on documenting work as a way of re-evaluating progress and re-prioritizing the remaining tasks. This insight highlights the opportunity for supporting the desired behavior by reinforcing the helpful behaviors (i.e., documentation).

4.2 Evaluation Study

The primary objective of the study is to examine if we can help individuals improve their work-nonwork balance through a reflection process driven by implementation intention and

mental contrasting (IIMC). We thus conducted a three-week between-participant study to address the following research questions:

RQ1 Does IIMC-based reflection on personal data improve behavior planning via enhanced understanding of obstacles on the way of work-nonwork balance and the opportunities to respond to these obstacles?

RQ2 Does IIMC-based reflection on personal data improve perceived work-nonwork balance via enhanced behavior plans during the study? If so, does the improvement surpass that of alternatives?

We asked study participants to create behavior plans to improve their work-nonwork balance and compared outcomes (Section 4.2.2) among four groups: (1) Data+IIMC (D+IIMC) group who used WoNoB's planning tool (Section 4.1.2) to draw insights from their data and create behavior plans within the IIMC framework, (2) IIMC group who were given standard IIMC instructions but did not reflect on their data to create their behavior plans, (3) Data-driven reflection (D-only) group who used the visual elements of the interface but were not given IIMC instructions in plan creation, (4) Basic control group who neither reflected on their data nor received IIMC instructions. Basic control and D-only groups were given instructions in positive thinking for creating their plans, which allowed us to control for the IIMC effect (i.e., we ensured the plans in these conditions are not based in IIMC, e.g., because of prior training). Moreover, positive thinking instructions are typically used in empirical studies of IIMC in comparison groups (e.g., [139]). Figure 4.8 illustrates the particulars and differences in the four conditions. Having these conditions allowed us to examine data review with and without IIMC instructions while also disentangling the contribution of data and IIMC to the outcomes.

Participants collected personal data during the first week of the study using the logging interface we had designed for active reporting of activities, locations, and progress toward goals (Section 4.1.1). They were then randomly assigned to one of the four conditions for creating behavior plans at the start of the second week of the study. They repeated the plan

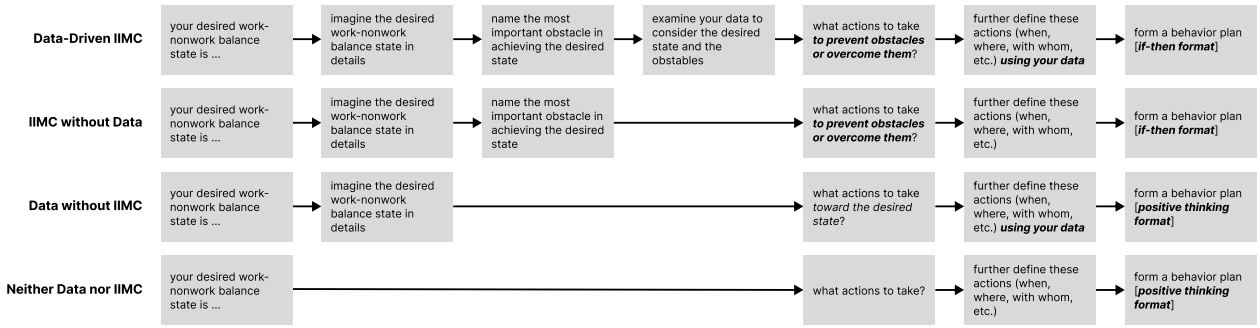


Figure 4.8: Intervention instructions given under different study conditions. The top row illustrates the instructions for data review with implementation intention and mental contrasting (D+IIMC condition). Data-related elements of the planning tool (i.e., filters, calendar, and summary) only appeared when data use was part of the instructions (i.e., these elements were only shown for D+IIMC and D-only conditions). If-then behavior plans were of the form "If ... (the obstacle or an opportunity to prevent it) arises, then I will do ... (actions in time, location, and other context) to overcome or prevent the obstacle." For example, "If I get emails after work hours, then I will use the focus time app on my phone to auto-hide it." Positive thinking plans were of the form "I want to ... (actions to take) to achieve my goal to ... (goal for improving work-nonwork balance)". For example, "I want to finish coding by 3pm to work on administrative tasks to achieve my goal to finish all work-related activities by 5pm."

creation at the start of the third week. Our analysis and reporting are based on data from the third week; the main purpose of the second week was to acclimate participants to planning in a given condition and to reduce the impact of the novelty of the experience on the outcomes. It additionally allowed participants to adjust logging to their needs for the plan creation of the third week. We detail the steps of the study in Section 4.2.3. Participants were compensated with $175 Amazon gift cards for completing the study activities. They received an additional $75 gift card if they fully logged their activities for 18 of 21 days of the study. Our study was approved by our institution's Institutional Review Board (IRB).

4.2.1 Participants

We reached out to a randomly sampled group of information workers at a global technology company where workers may collaborate across multiple timezones. Our recruitment email

advertised the study as an investigation into different ways of improving work-nonwork balance using tools that allow individuals to create and protect time for activities that matter to them. It also included a brief screening survey, which we used to only enroll employees who resided in the US and were able to commit to the logistics of the study, those who met the technology requirements for participation, and those who met the requirements for benefiting from the solutions we offered.

Specifically, we screened respondents to our call to ensure participants were available for the duration of the study and anticipated that their schedules during the study would be representative of their typical schedules. The former requirement guaranteed the completion of study activities as planned, while the latter was important to control for extremes of overload (e.g., a major deadline) or underload (e.g., traveling or personal leave) as the study content and data were less relevant under these extremes. The primary technology requirement was to have the Microsoft Teams mobile application installed to allow participants to receive and engage with daily reminders during both working and nonworking hours.

We drew from behavior change literature to limit participation to people who report not having work-nonwork balance (i.e., they have recognized the problem area to address) and are in the preparation stage of change (i.e., they are *ready* for change but have not taken any actions yet). This group of people has the intent for change and can substantially benefit from support in doing so [155]. Support may be useful at later stages of change too but we scoped our research to this particular stage for experimental control purposes: we chose to focus on a stage of change where the affordances of our tools are well suited for addressing the known challenges, in the absence of evidence on whether and how our approach leads to improved outcomes. Behavior discovery (i.e., identifying the relevant actions to take) is a key but challenging step of the preparation stage [66] and our tool offers multiple avenues for addressing this challenge (see walk-through of Section 4.1.2 for details).

Our process led us to recruit 48 adult participants (12 per condition). From these, we removed 5 participants (2 in basic control, 1 in D-only, and 2 in D+IIMC conditions) over the course of the study because of personal emergencies (e.g., travel due to family health concern)

Table 4.1: Demographics of Study Participants. Participant ID assignments in each condition are given in the last row and will be used in quoting participants.

	Total	Basic control	D-only	IIMC	D+IIMC
Man	22	5	6	7	4
Woman	20	5	5	5	5
Unspecified Gender	1				1
18-25 years old	1			1	
26-35 years old	11	3	3	4	1
36-45 years old	13	4	2	5	2
46-55 years old	15	3	6	1	5
56-65 years old	2			1	1
Unspecified Age	1				1
Graduate degree	20	6	2	6	6
Post-Secondary degree	22	4	8	6	4
Unspecified Education	1		1		
Individual Contributor	19	6	2	4	7
Manager	21	4	6	8	3
Unspecified Job Role	3		3		
Range of Participant IDs	1-43	34-43	23-33	11-22	1-10

or low compliance. We report findings for the 43 remaining participants. Table 4.1 summarizes demographics for the study and across four study groups. Briefly, 21 identified as male, 21 as female (1 preferred not to identify their gender). 12 were 35 years and younger, while 20 were 36 years and above (1 participant did not provide their age). All participants who provided their education information (all but 1) had post-secondary degrees of which 20 also had a graduate degree. They occupied both management (21) and individual contributor (19) roles (3 did not provide roles).

4.2.2 Measures

We collected information on the relevant outcomes along with factors that can affect the outcomes in addition to the experimental condition. We also collected information to verify that WoNoB's planning tool successfully supports a reflective process as the key target of IIMC instructions.

Addressing **RQ1** (if we can improve understanding of obstacles and opportunities for action), we asked participants for ratings on whether the planning tool helped them:

1. Find obstacles that get in the way of following their plans (obstacle identification)
2. Find problematic behaviors in achieving their plans (problematic behavior identification)
3. Identify opportunities for actions in the short-term, i.e., within a week (short-term action identification)
4. Identify opportunities for actions in the long-term, i.e., beyond a week (long-term action identification)
5. Identify opportunities for rescheduling plans (rescheduling ability)
6. Become more aware of their adherence to plans (adherence awareness)
7. Become more determined in adhering to their plans (adherence determination)

We obtained these ratings on a 5-point Likert scale (1: strongly disagree – 5: strongly agree) at the end of week 3 of the study as doing so earlier could have confounded the natural use of the planning tools (planning tools varied based on experimental condition; see further details in Section 4.2.3).

Addressing **RQ2** (if we can improve work-nonwork balance), we considered perceived work-nonwork balance and time-management at baseline (i.e., start of week 1 of the study) and exit (i.e., at the end of week 3 of the study). We asked participants to provide this information based on their experience over the prior week at each time point. Specifically, we obtained involvement, effectiveness, and affective balance subscales as measures of perceived work-nonwork balance [193]. We derived a measure of time-management from [139] as the sum of Likert ratings (1: Never – 5: Very Often) of the following statements: feeling pressed for time, managing time easily, keeping planned times (e.g., appointments, meetings, blocked time) easily. The first item was reverse-coded before obtaining the sum.

We obtained measurements for a number of factors that could influence the outcomes in addition to the experimental condition. Doing so allowed us to either establish that conditions

were comparable with respect to a factor or to control for the effect of the factor in the comparisons. The factors we measured pertain to personal characteristics that impact the effectiveness of different aspects of IIMC-based reflection, individual contexts, and activities. With respect to personal characteristics, we collected measures of self-efficacy [27], socially prescribed perfectionism [86], and conscientiousness [72], as past research has shown they influence the effectiveness of IIMC use. We also collected reflective tendencies [148] that can affect whether individuals are inclined to engage in a reflective process. We obtained these measures at the beginning of the study. With respect to individual context, we asked about levels of work and nonwork load as well as caregiving responsibilities and resources at study onboarding. With respect to individual activities, we obtained weekly measures of the insight subscale of technology-supported reflective inventory (TSRI-insight) [13] to measure the level of reflection associated with using the logging interface, as past research indicates the act of recording information is reflective in and of itself [195]. We consider the reports of TSRI-insight for the last week of logging as the most stable representation of reflection at logging.

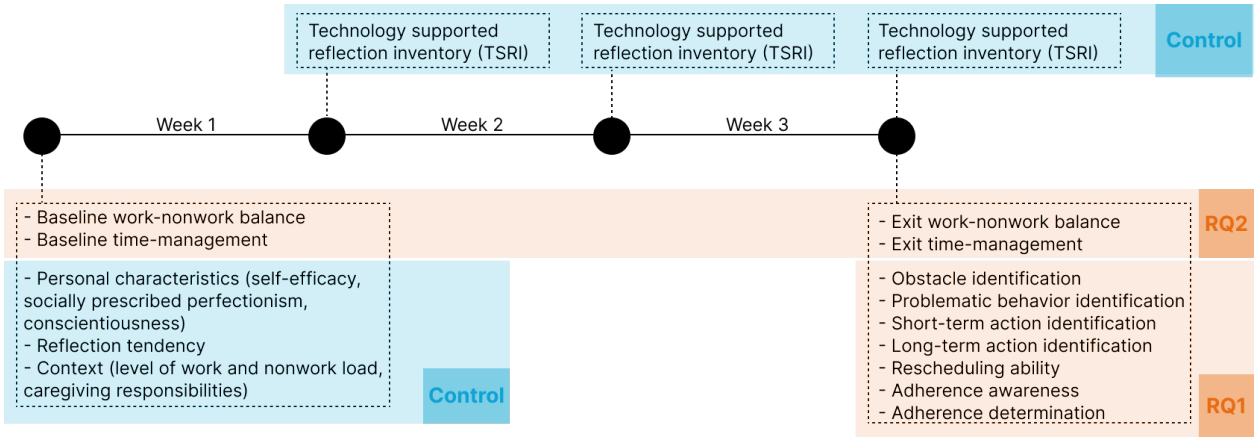


Figure 4.9: Study measurements. Measurements addressing **RQ1** were obtained at the end of week 3 of the study. Baseline and exit measurements of **RQ2** were obtained at the beginning and at the end of the study. Several control variables including personal characteristics, reflection tendency, and context were obtained at the beginning of the study. Weekly measures of planning and logging reflection were also obtained as control via insight subscale of technology supported reflection inventory (TSRI).

We obtained a weekly measure of TSRI-insight with respect to the planning tool to verify that it successfully supported reflection and helped participants gain insights into achieving their work-nonwork goals. We use reports of reflection on planning tools in week 3 as the representation of planning reflection. Figure 4.9 details the measurements and their timing.

4.2.3 Procedure

Participants went through several onboarding activities upon enrolling in the study. They completed standard questionnaires on their perceived self-efficacy [27], socially prescribed perfectionism [86], conscientiousness [72], and reflective attitude [148]. They also responded to questions about their work and nonwork load as well as caregiving responsibilities and resources, described their views of the desired work-nonwork balance, and customized the list of activities and locations to log.

At the start of each week of the study, participants reported their perceived work-nonwork balance and time-management over the past week. They then reviewed their work-nonwork balance goals and the plans to achieve those goals, updated their goals as needed, and modified/refined their behavior plans except at week 1; participants described work-nonwork goals and plans for the first time at the onset of week 1. All groups used the basic control tool for behavior planning of week 1 as a baseline. Starting at week 2, participants used their assigned tools for behavior planning at the start of the work week. That is, D+IIMC and D-only groups used the planning tool in Figure 4.2 with respective instructions on the left per Figure 4.8, whereas IIMC and basic control groups received their respective instructions without having access to the visual elements of the planning tool (i.e., filters, calendar, or summary parts). We displayed each participant their description of the desired work-nonwork balance at the time of goal setting and behavior planning to ensure their ideals are available for consideration in decision-making. This strategy helped participants come up with plans that were more relevant to them. Participants responded to TSRI-insight after using the planning tool.

During the week, participants were reminded randomly five times a day during waking hours to log their activities, whereabouts, and progress. We asked participants to complete TSRI-insight

with respect to the logging tool at the end of the week.

At the end of week 3 of the study, participants reported their perceived work-nonwork balance and time-management over the last week of the study. They then rated the statements that we specifically designed to address **RQ1** (see the details in Section 4.2.2). They also responded to open-ended questions about their experience in the study. These questions asked about what they liked/disliked, how they used the tools, and ways we can improve them.

4.2.4 Analysis

We created regression models of the form $outcome \sim data + iimc + iimc \times data + control$ to address **RQ1**. We created a separate model for each measure of **RQ1** as $outcome$. We used $data$ and $iimc$ as binary variables that captured whether reflection on data and IIMC instructions were received. Therefore, both $data$ and $iimc$ were 1 for D+IIMC group, whereas they were both 0 for basic control group. Only $data$ was 1 for D-only group, while only $iimc$ was 1 for IIMC group. We considered two categories of variables to select the $control$ variables: (1) demographics (Table 4.1) as past research has shown differences in outcomes in relation to these characteristics [78, 57], and (2) personal characteristics as well as individual context and activities (see Section 4.2.2). We included as the $control$ those variables that differed among conditions or showed significant correlations with the outcome of interest. No demographic variables were included in the models as Fisher's exact test showed no significant differences across conditions on demographics (Fisher's test is preferred over χ^2 test for small samples). We included self-efficacy in the model of determination for plan adherence given their significant correlation. Similarly, we included conscientiousness in models for identification of obstacles and rescheduling opportunities. We controlled for levels of reflection at logging in all models as this variable was significantly correlated with all outcomes. Control variables included for each outcome are listed in Table 4.4. We used the lm function of `stats` package in R [156] to create the regression models.

Addressing **RQ2**, we used mixed ANOVA with time (baseline to exit) as the within factor and condition as the between factor to examine if mean changes in perceived work-nonwork

balance subscales and time-management differ across conditions. We did a follow-up analysis on significant main effects when doing so was possible (i.e., there was no interaction effect between time and condition): dependent-sample t-test to follow up on the main effect of time and Tukey's HSD to follow up on the main effect of condition. We used the `rstatix` package [97] to test ANOVA assumptions and obtain test statistics for mixed ANOVA analysis and Tukey's HSD. We used the `stats` package for paired-sample t-test [156].

We took several steps in verifying various aspects of our methodology. First, we examined the relevance of the custom measures of **RQ1** to the work-nonwork balance construct by examining the correlations between these custom measures and the validated measure of perceived work-nonwork balance [193]. If correlated, we have evidence that the use of the planning tool has activated mechanisms intended by IIMC in the context of work-nonwork balance. Second, we used a one-sided t-test to compare week 3 measurements of TSRI-insight on planning tools with the mid score of 12 on this subscale for each study condition to check whether planning tools successfully supported reflection. This strengthens the assumption that reflection processes we assumed to be active are in fact active. We applied Bonferroni correction to the nominal α of 0.05 (i.e., we used p-value of 0.0125 to establish significance) to account for type I error of multiple comparisons.

We triangulated the quantitative analysis of our measures with the qualitative analysis of survey responses as the latter could contextualize quantitative patterns. Triangulation is commonly used in HCI as a way of obtaining a more "reliable, holistic and well-motivated understanding of phenomena" [149]. Our qualitative analysis drew upon reflexive thematic analysis methodology [21], where we considered comments in light of the outcomes we were examining quantitatively: I reviewed comments in relation to such objectives as change in awareness, planning process, or perceived improvements, while iteratively coding for additional nuances (e.g., factors influencing planning). A collaborator reviewed my coding of comments. Coding disagreements were resolved through discussion among co-authors.

4.3 Results

We find that the planning tools that incorporate instructions within the framework of implementation intention and mental contrasting (IIMC) successfully guide reflection in behavior planning. We also observe the importance of IIMC instructions in data review for increased awareness of plan adherence and the ability for rescheduling, qualities that are important in realizing work-nonwork balance goals. Moreover, our observations indicate significant improvements in perceived work-nonwork balance and time-management across experimental conditions. We detail our findings in the following subsections. Our use of the word 'significant' in the reports means 'statistically significant'. We use the standard significance level of 0.05 unless adjustments were needed (see Section 4.2.4).

4.3.1 IIMC Instructions Guide Reflection in Behavior Planning

Comparing TSRI-insight scores for week 3 of planning across conditions, we find that D+IIMC and IIMC groups report levels of reflection that are significantly larger than the mid-point score of 12 on the TSRI-insight subscale for their assigned planning tool (D+IIMC mean=15.6 and IIMC mean=16.25; Table 4.2). We did not find statistically significant evidence that D-only and basic control groups report levels of reflection above the mid-point threshold (D-only mean=12.73, basic control mean=14.90). Figure 4.10 provides the distribution of scores across conditions. We repeated the analysis with a one-sided Wilcox signed rank test for D-only and basic control groups given the non-normal distributions of scores for these conditions. We did not find significant evidence that the level of reflection in these groups surpasses the mid-point threshold.

4.3.2 Increased Plan Awareness and Rescheduling for Work-Nonwork Balance in D+IIMC

Custom-defined measures of **RQ1** are significantly correlated with perceived work-nonwork balance subscales. Correlation coefficients are moderate to large (min = 0.32, max = 0.68; Table 4.3). We thus have evidence for the relevance of the measures and can more confidently interpret them.

Table 4.2: Descriptive statistics of TSRI-insight scores for week 3 of planning across conditions along with one-sided t-test comparisons against the mid-point score of 12. TSRI-insight values vary in the range of 3-21 (larger values indicate higher levels of reflection). The significance level is coded with stars (*: p < 0.05, **: p < 0.01).

Group	Mean	Std.	t-test	p
D+IIMC	15.60	3.69	3.09	**
IIMC	16.25	4.29	3.43	**
D-only	12.73	5.20	0.46	
Basic control	14.90	4.95	1.85	

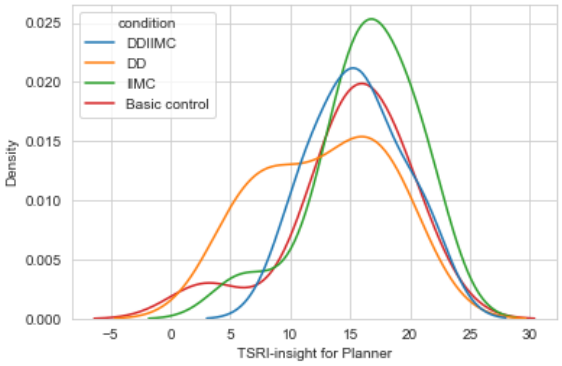


Figure 4.10: Distribution of TSRI-insight scores for planning interface in week 3 for D+IIMC (blue), IIMC (green), D-only (orange), and basic control (red) conditions.

We find affirmative support for **RQ1**. We observe significant differences across conditions for measures of rescheduling ability, adherence awareness, and adherence determination: participants in D+IIMC condition report significantly higher ratings for the benefit of their planning tool in enabling them to reschedule ($F(1)$=4.74, p=0.036) and to become more aware of their adherence to the plans they have for improving work-nonwork balance ($F(1)$=4.11, p=0.050). However, D-only group's determination for adhering to their plans is significantly diminished compared to other groups ($F(1)$=5.73, p=0.022). Table 4.4 lists the regression coefficients across input variables for different outcomes.

D+IIMC participants' survey responses qualitatively contextualized our observations above. Some participants explicitly attributed the increased awareness of their plan compliance to IIMC-based data review. For example, P2 said *"I reviewed my study behavior plan and last week saw blocks where I did not follow the nonwork and work plan."* Similarly, P7 said she most liked *"the visibility into my own schedule and how closely (or not) I was able to stick to my calendar."* Describing her use of data toward rescheduling, P9 said *"I could see that there is time gaps in my day that allow me to think about moving my blocked time."*

Table 4.3: Correlations coefficients between custom measures of **RQ1** and standard measures of work-nonwork balance [193]. The significance level is coded with stars (*: $p < 0.05$, **: $p < 0.01$).

	Involvement balance	**Affective balance**	**Effectiveness balance**
Obstacle identification	0.52 **	0.51 **	0.57 **
Problematic behavior identification	0.63 **	0.48 **	0.6 **
Short-term action identification	0.48 **	0.32 *	0.48 **
Long-term action identification	0.52 **	0.39 *	0.47 **
Rescheduling ability	0.4 **	0.52 **	0.45 **
Adherence awareness	0.47 **	0.52 **	0.55 **
Adherence determination	0.48 **	0.47 **	0.49 **

We find that levels of reflection at logging (measured using weekly TSRI-insight scores) are significantly related to a better understanding of obstacles and opportunities: the more participants reflected at the time of logging, the more they could identify obstacles or problematic behaviors when planning ($F(1) = 26.48, p \ll 0.001$). Similarly, they were better at finding opportunities for action in the short and long-term or options for rescheduling ($F(1) = 16.77$, 20.27, and 29.17, respectively, $p \ll 0.001$). Moreover, they felt more aware of their plan adherence and more determined in following through with their plans ($F(1) = 4.34$ and 32.67, respectively, $p < 0.044$). Conscientiousness significantly explained additional variations in identifying opportunities for rescheduling ($F(1) = 6.41, p = 0.016$).

The value of logging was also evident in participant responses to open-ended survey questions. For example, P34 said *"Putting down all my tasks for the day and the time I did them allowed me to reflect on which areas I can improve on."* Or P24 said *"Having to log all my time was a real wake-up-call to how my week passed."*

In addition to an increased awareness and understanding of obstacles, we find that D+IIMC instructions (Figure 4.3) guided participants to create a concrete behavior plan that directly addressed the obstacle identified through data. For example, P3 was made aware that responding to emails after work hours was a norm, not an exception: *"I see that I was working from 8 pm*

Table 4.4: Coefficients for regression models of **RQ1**. Outcomes appear on the rows. We report estimated coefficients for all input variables that were included for each outcome along with the standard error for the estimates in parenthesis (an empty cell means the variable was not part of the model). The range of values for logger reflection, self-efficacy, and conscientiousness are 3-21, 1-5, and 5-50 respectively. Larger values indicate higher levels of reflection, self-efficacy, and conscientiousness. Significance level is coded with stars (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

	data	**iimc**	**data × iimc**	**logger reflection**	**self-efficacy**	**conscientiousness**
Obstacle identification	0.08 (0.36)	0.06 (0.36)	0.31 (0.51)	0.14*** (0.03)		0.04 (0.04)
Problematic behavior identification	0.48 (0.35)	0.43 (0.34)	-0.36 (0.49)	0.14*** (0.03)		
Short-term action identification	-0.07 (0.44)	-0.00 (0.43)	0.13 (0.62)	0.13*** (0.03)		
Long-term action identification	0.19 (0.33)	0.24 (0.32)	-0.05 (0.46)	0.11*** (0.02)		
Rescheduling ability	-0.45 (0.38)	0.23 (0.38)	1.16* (0.53)	0.15*** (0.03)		0.06* (0.03)
Adherence awareness	-0.80* (0.36)	-.0.57 (0.35)	1.02* (0.50)	0.06* (0.03)		
Adherence determination	-0.84* (0.35)	-0.24 (0.34)	0.84 (0.52)	0.15*** (0.03)	0.53 (0.31)	

to 9.30 pm last week from the app." He then identified phone notifications as an obstacle: *"I get notifications on my phone when someone emails me on my work email. If it's important or interesting, I tend to answer those."* Finally, he created a behavior plan as if-then statements that addressed his observation: *"If I get emails after work hours, then I will use focus time app on my phone to auto-hide it to overcome or prevent the obstacle."*

4.3.3 Increased Perceived Work-Nonwork Balance and Time-Management during Study

Our data positively supports **RQ2** with respect to improvements in work-nonwork balance as we see enhanced measures over the study duration across all study conditions. However, there is no evidence that D+IIMC improvements are surpassing the alternatives. Mixed ANOVA models pertaining to **RQ2** indicate no significant interaction effect between time and condition. We can thus directly interpret the main effects. We find a significant main effect for time: participants more favorably rate their time-management at exit compared to baseline with average improvements of 1.84 over the range of 3 to 15 for this measure (Figure 4.11; $t(42) = 5.13, p \ll 0.001$). Moreover, participants report large to very large improvements in

measures of perceived work-nonwork balance from baseline to exit. Specifically, affective, effectiveness, and involvement balance scores improve by 0.53, 0.55, and 0.98 on average over the range of 1 to 5 for these measures (Figure 4.12; $t(42) = 4.01, 4.08$, and 5.95, respectively, $p \ll 0.001$ for all).

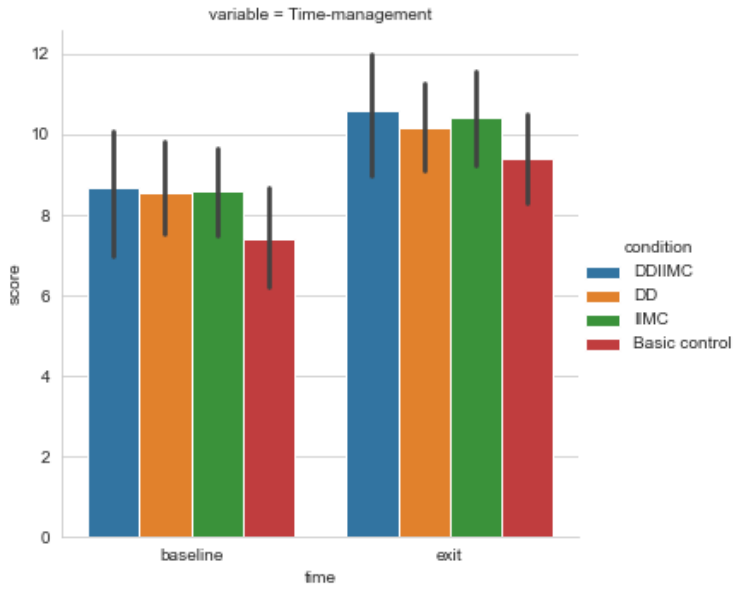


Figure 4.11: Baseline and exit scores of time-management. D+IIMC, D-only, IIMC and basic control conditions are color-coded by blue, orange, green, and red respectively. Error bars indicate a 95% confidence interval based on group standard deviations. There is a very large but condition-agnostic improvement.

However, there is no significant main effect for condition: the average change over time does not differ across conditions – work-nonwork balance and time-management similarly improved for all conditions, independent of the planning interface they used over the course of the study. Table 4.5 provides descriptive and test statistics for the measures. Qualitative survey responses corroborated the observed improvements in work-nonwork balance and time-management. They also brought to light important nuances in the multi-stakeholder context of work and nonwork. P18 explicitly connected improvements in work-nonwork balance to their participation: *"The exercise helped me prioritize my nonwork activities in my day-to-day life and achieve better balance"*. P37 made a similar comment, while acknowledging the tension between work and nonwork and her agency in making choices and acting on them: *"This process made me very mindful of work and nonwork times and [I] was more deliberate about my choices (when I had choice) [to] step away both in the mornings to quickly check my messages or the evenings to get one last thing out the door, and many times I do have the choice but not every time."*

Some participants described the realization of the role of external influence in their work-nonwork balance as the most helpful aspect of their participation, confirming that work-nonwork balance does not pertain only to the individual but the people and work context around them [19]: *"Recognition that I don't have as much agency in what makes me busy at work. I'm at the whims of other people's calendar and requests that can disrupt my best laid plans. This is likely driven from my desire for impact, recognition, and eventually promotion but the questions prompted during this study really surfaced this"* (P40). Some took this recognition to action and wanted to share their data-backed insights with others (peers, managers, family members) to mitigate the external demands. P4 suggested *"a way to collaborate with your life partner / spouse if you are sharing home responsibilities"* as the main improvement to the tool.

Condition-independent improvements despite the differences in planning-related measures of **RQ1** suggest that a factor other than planning drives the improvements in perceived work-nonwork balance and time-management. To further examine the factors that influence the changes in perceived work-nonwork balance and time-management, we built regression models of the form $\Delta outcome \sim data + iimc + data \times iimc + other$. $\Delta outcome$ is the difference in exit with respect to baseline for $outcome$ scores; $data$ and $iimc$ were defined similar to regression models of **RQ1**. Similarly, $other$ was selected from personal characteristics or individual activities (i.e., logging) variables based on whether these variables were significantly correlated with the outcomes. The selections were: self-efficacy for all subscales of perceived work-nonwork balance, conscientiousness for all subscales but the involvement balance, and levels of reflection at logging for all measures of perceived work-nonwork balance and time-management.

Reflection at logging turned out to be a significant predictor of change in all work-nonwork balance and time-management outcomes ($F(1) = 12.96, 9.41, 5.47$, and 8.01 for time-management, affective, effectiveness, and involvement balance, respectively, all $p < 0.025$), similar to earlier observations that it was significantly related to plan awareness and rescheduling. Self-efficacy and conscientiousness did not significantly explain variations in any of their respective models (Table 4.6). Despite the benefits of enabling reflection and prioritizing

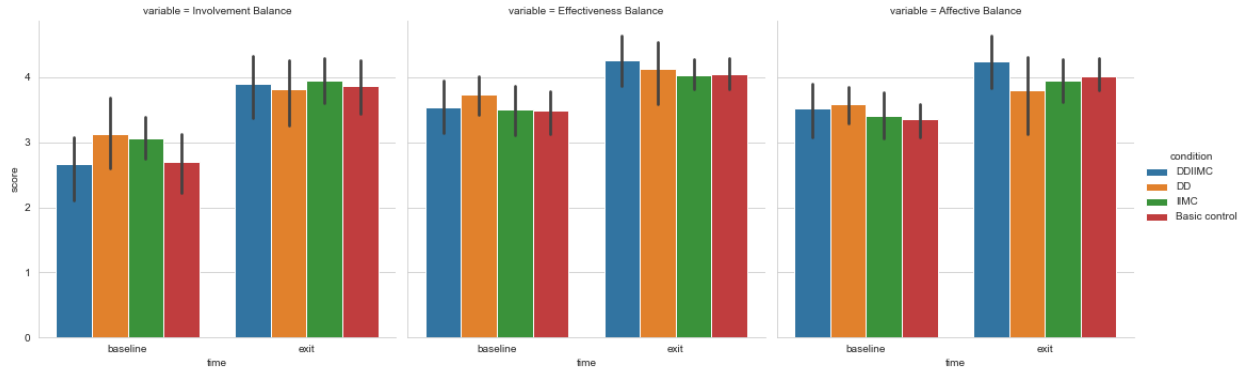


Figure 4.12: Baseline and exit scores of perceived work-nonwork balance. From left to right, we have involvement balance, effectiveness balance, and affective balance. D+IIMC, D-only, IIMC and basic control conditions are color-coded by blue, orange, green, and red respectively. Error bars indicate a 95% confidence interval based on group standard deviations. There are large to very large improvements across all subscales of perceived work-nonwork balance but the improvements are not condition dependent.

work-nonwork balance, participants also hinted at the burden of data collection. For example, P29 said *"I like the daily reminders although logging time does cause some stress to get it done. That said, the reminders that come in Teams remind me to make this (referring to her nonwork goal) a priority."*

4.4 Discussion

We examined whether and how the use of implementation intention and mental contrasting (IIMC) to scaffold reflection on personal data can help decision-making in creating behavior plans for improving work-nonwork balance goals. Our results demonstrate the value of IIMC-based data review. More importantly, they highlight additional opportunities for supporting decision-making with data. Furthermore, we found that reflection helped participants become more aware of their agency under external forces in the multi-stakeholder context of work and nonwork, where data can play a boundary-negotiating role [36]. Although our study is in the context of work-nonwork balance, several findings and their implications have broader applicability which we further discuss below. In the following subsections, we

Table 4.5: Descriptive and F statistics for variables of **RQ2**. Subscales of perceived work-nonwork balance vary in the 1-5 range. Time-management scores are in the range of 3-15. Larger values indicate better work-nonwork balance and time-management.

	Affective balance				**Effectiveness balance**				**Involvement balance**				**Time-management**			
	Baseline		**Exit**		**Baseline**		**Exit**		**Baseline**		**Exit**		**Baseline**		**Exit**	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
D+IIMC	3.52	0.71	4.24	0.69	3.54	0.71	4.26	0.68	3.43	0.57	3.87	1.26	8.7	2.67	10.6	2.55
IIMC	3.4	0.64	3.95	0.59	3.5	0.71	4.03	0.45	2.94	0.99	3.56	0.74	8.58	1.98	10.4	2.11
D-only	3.58	0.49	3.8	1.07	3.73	0.53	4.13	0.87	3.48	1.1	4.03	0.69	8.54	2.02	10.2	2.04
Basic control	3.36	0.43	4.02	0.43	3.48	0.56	4.04	0.42	3.4	0.84	4.07	0.61	7.4	2.12	9.4	1.84
$F_{condition}(3, 39)$			$0.40, p = 0.76$				$0.43, p = 0.74$				$1.27, p = 0.3$				$1, p = 0.4$	

Table 4.6: Coefficients for regression models of **RQ2**. Outcomes appear on the rows. Δ is the change from baseline to exit. We report estimated coefficients for all input variables that were included for each outcome along with the standard error for the estimates in parenthesis (an empty cell means the variable was not part of the model). The range of values for logger reflection, self-efficacy, and conscientiousness are 3-21, 1-5, and 5-50 respectively. Larger values indicate higher levels of reflection, self-efficacy, and conscientiousness. Significance level is coded with stars (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

	data	**iimc**	**data $\times$ iimc**	**logger reflection**	**self-efficacy**	**conscientiousness**
Δ Affective balance	-20.3 (0.36)	-0.19 (0.36)	0.36 (0.52)	0.08** (0.03)	0.34 (0.32)	0.00 (0.03)
Δ Effectiveness balance	-1.11 (0.39)	-0.10 (0.38)	0.27 (0.57)	0.07* (0.03)	0.03 (0.35)	0.02 (0.03)
Δ Involvement balance	-0.43 (0.45)	-0.34 (0.44)	0.64 (0.67)	0.09** (0.03)	0.22 (0.40)	
Δ Time-management	-0.38 (0.93)	-0.27 (0.91)	0.28 (1.30)	0.25*** (0.07)		

more closely interpret our results, discuss the opportunities for further incorporating IIMC in planning with personal data, share ideas for bringing IIMC to the workplace, and critically examine the limitations of our work.

4.4.1 IIMC-Based Data Review Supports Behavior Planning but There Is More to Goal Realization

We observed increased subjective awareness of adherence to goals and capability for rescheduling plans among D+IIMC participants in our study (Section 4.3.2). These observations provide evidence that IIMC scaffolds personal data review in ways that support better planning

for work-nonwork balance goals. We note that planning reflection levels for D+IIMC and IIMC groups surpassed the mid-point threshold while the same did not happen for D-only and basic control groups (Section 4.3.1). The case for D-only is particularly interesting. In the absence of guidance when reviewing data there is no gain but loss: participants in D-only report diminished determination in following their plans (Section 4.3.2). This can be explained through the interrelation between motivation and IIMC [138]. Reviewing data likely brought to light the gap between the desired and actual state of work-nonwork balance. This realization subsequently undermined participant determination as there was no guidance on addressing the gap. Our study shows that the IIMC technique explicitly supports people in overcoming the gap.

We also found condition-independent improvements in perceived work-nonwork and time-management from baseline to exit (Section 4.3.3). Study procedures involved multiple behavior change techniques, including goal-setting, tracking (in the form of logging), and behavior planning. While we isolated and varied the latter only, other techniques could drive the improvements individually or in combination with the rest. We showed that reflection associated with logging was a significant predictor of improvements above and beyond planning (Section 4.3.3).

4.4.2 IIMC-Guided Micro-Reflection on Data in Recent Past Can Support Behavior Planning

Observations of increased rescheduling ability and adherence awareness without loss in adherence determination in Section 4.3.2 demonstrate the importance of guided *macro-reflection* where people review their personal data to identify opportunities for action within the IIMC framework. However, such macro-reflection was not the only source of benefits for our participants. Logging data and reflecting on one's behavior when doing so were key to success in planning-related activities (Section 4.3.2) as well as improvements in perceived work-nonwork balance and time-management (Section 4.3.3). These observations are consistent with other reports (e.g., [112, 31]), which similarly found reflection happens at the time of logging. In our setting, reflection at logging is a form of *micro-reflection* where people pay attention to their behaviors in the very recent past and within a short window of time when

logging. In doing so, they notice divergence from desired behaviors, the underlying reasons (i.e., obstacles), and potential solutions. This is evident in the significant positive relations between levels of reflection at logging and ratings of the ability to identify obstacles, problematic behaviors, and opportunities for action (Section 4.3.2). It is noteworthy that macro-reflection on data when planning did not seem to explain these outcomes; the D+IIMC group did no better than the others. Not surprisingly, micro-reflection at logging also showed the importance for rescheduling and awareness of plan adherence and thus complemented macro-reflection at D+IIMC planning (Section 4.3.2). In light of these results, we call out for attention to and investment in leveraging IIMC for micro-reflection. The ideas and lessons from research on just-in-time adaptive interventions can be leveraged (e.g., in determining the number, timing, and content of IIMC-based micro-reflection prompts [113, 114, 136, 132]).

We contrast our position with much of the past HCI research on implementation intentions, which has focused on the *automatic detection of identified situations* for triggering action (e.g., [16, 152]). Contrary to previous work, we emphasize micro-reflections through *low-cost and simple logging for identifying situations*. With IIMC guidance, micro-reflection helps people identify relevant situations for action. Macro-reflections then support specifics of actions to take.

4.4.3 Incorporate IIMC Scaffolding in Existing and Future Workplace Tools

Our results indicate the value of incorporating IIMC scaffolding in workplace well-being context through both micro- and macro-reflection techniques. There are several possibilities for supporting these techniques. With respect to micro-reflection, the simplest practical implementation can use fixed reminders employees set for themselves to think back on their activities, whether those activities align with their work-nonwork balance goals, reasons for divergence, and ways to overcome or prevent it. This is similar to the reflection feature of such apps as Viva[2] but underscores explicit scaffolding of the reflective process. Moreover, calendar or meeting/communication applications can prompt quick reflection on how well the time was

[2]https://learn.microsoft.com/en-us/viva/insights/personal/teams/viva-insights-reflect

spent after each scheduled block of time or bout of communication. Predicting opportune moments for micro-reflection can further empower employees to more easily integrate micro-reflection in their day-to-day work (e.g., predicting moments of transition or break similar to [100] or when the chances of engagement are high as in [150]). However, this is a delicate design space as competing factors may complicate micro-reflection at opportune moments as noted in other research [110] (e.g., momentary negative affect from micro-reflection can adversely influence engagement).

With respect to supporting macro-reflection, the simplest and most available solution for employees is to designate a time to review their calendars, clarify their goals, reflect on existing obstacles and problematic behaviors, and identify opportunities for action. There is already evidence for productivity and well-being benefits of reflective goal-setting (e.g., in the context of information work [129]). Our emphasis is on the value of IIMC instructions in guiding the reflective process with personal data, and we specifically advocate for extending the same strategy to both work and nonwork goals. We acknowledge that standard calendar data may not be as rich as the kind of data underlying our findings. Nonetheless, it may be a useful resource and should be studied within a larger inquiry into the minimum requirements for the level of detail, coverage, and granularity (e.g., 30-minute vs. hour-long intervals) of personal data. In another approach, it is possible to bring some personalized automation to the process where a system such as Pearl [91] can assist in the sense-making process of personal data to help identify potential obstacles and opportunities for action as employees define relevant concepts for the system to help them more closely analyze their behaviors. This approach heavily relies on the availability of telemetry and other passively-sensed data to employees, although we acknowledge that there are important privacy and ethical concerns with such data that should be separately addressed. It is best to combine micro- and macro-reflection, independently of the specifics of implementations of each.

The focus of our exploratory experiment was on improving individual-level decision-making and action toward bettering work-nonwork balance. However, it is too simplistic to assume work-nonwork balance, or any other workplace well-being topic is individually scoped. Obstacles may be driven by external factors. There might be a need for communicating and

coordinating actions with others (e.g., direct reports, peers, supervisors, or family members). Personal data and IIMC-based insights can act as boundary-negotiating artifacts in navigating, communicating, and coordinating with others, albeit there are nuances around potentially conflicting employee vs. organization goals as well as power dynamics. Future research should closely examine D+IIMC within a group context. More broadly, it is important to explore ways IIMC-guided reflection on personal data can support change at the organizational and social level. The increased awareness and recognition of external factors that result from the reflection can be key ingredients of individual's agency for change, as we observed among participants who started conversations with their managers (Section 4.3.3).

4.4.4 Limitations

Our study relied on rich but manually logged activity, location, and progress information. Such data cannot be assumed in most practical setting. While this assumption helped narrow down the confounding factors, it is important to reproduce our results with more realistic data, such as calendar information or automatically sensed data using all-purpose or personalized detectors. As we later found, logging was a critical contributor to the improvements we observed. Reproducing our study with passively-sensed data can further tease apart the specific value of IIMC-based macro-planning. We considered three sources of information but our analysis did not allow us to examine the relative importance of these different sources for the effects we observed. Moreover, we did not collect information on any additional sources of data participants might have found useful. We cannot thus speak to specific data requirements for a system such as ours. Our analysis primarily relied on subjective self-reports given the difficulties in defining generic objective measures of success for short-term goal realization across a variety of goals that our participants defined over the course of the study. We also note that our study was a field deployment and we did not have control over participant interactions with our tools and whether they leveraged other interfaces. For example, we could not prevent basic control or IIMC groups from consulting with their calendar as they were creating their behavior plans even though we assumed participants in these conditions did not rely on such data. We restricted participation to

people in the preparation stage of behavior change. Future work can study whether IIMC-based data review in other stages of change.

4.5 Contributions to Thesis

Goal-directed personal informatics involves different self-tracking and behavior change activities. This chapter explored enabling one such activity: goal-aligned action planning in the complex context of improving work-nonwork balance where many individuals struggle with forming robust behavior plans. Specifically, implementation intention and mental contrasting were used as techniques to guide reflection on personal data in action planning. These techniques help individuals review their data to identify not only the obstacles on the way of achieving work-nonwork balance, but also opportunities to address them. Our findings indicate the benefits of this scaffolding technique for reflection and underscores the need for exploring other such techniques in goal-directed designs.

Chapter 5

GOAL-DIRECTED ANALYSIS OF PERSONAL DATA

Chapter 3 showcased diverse and personalized tracking afforded by goal-directed designs. In this chapter, I consider the new requirements for analysis posed by the personalized and goal-directed data due to its high between and within variations: different people record different information, record the same information differently, or desire different uses of the same information. One person may change their recording or use of information over time. Specifically, high flexibility of goal-directed designs allows people to have different tracking goals (Section 3.3.5). For example, one person recorded numbness (PT01), while another recorded clumsiness (PT08). Screen time as a potential contributor mattered to one (e.g., PT02 or PT04), whereas high altitudes mattered to another (PT03). High flexibility also means recording or using the same type of information differently. Figure 5.1 illustrates some of the ways participants reported migraine duration: daily length in hours (PT03), time range in one column (PT08, PT10), or timing and duration tentatively recorded as free-form text (PT09). These reports appeared under varied columns, depending on individual preferences. As an example of high variability in the desired use of the information, some participants cared more about the length of migraine episodes (i.e., duration over consecutive days of migraine) rather than the daily length of migraine (PT01, PT04). Goal-evolution drives within variations as one person changes their recording or use of information over time (Figure 5.2). The change was sometimes manifested as records under a new column (PT10 - migraine duration) or a set of new columns (PT04 - location of migraine). Other times recording changed within one column, with a new data type (PT07 - sleep, PT08 - nausea) or new semantics for the same data type (PT06 - migraine presence). The latter was most common in evolving the tracking model from daily to event-based and vice versa where the semantics of blank entries changed.

PT03

Migraine Duration
2
4
14
0
0

PT08

Migraine
11:45-14:14
11:50-xx:xx
14:40-xx:xx

PT10

Start and End Times
23:00-23:30
20:30-xx:xx
19:30-22:30
12:00-xx:xx
06:00-09:00

PT09

Migraine Duration
overnight
12pm to nighttime
late night into morning
carried over from last night, felt mild in the morning, stayed until afternoon getting more intense
4+ hours
night time maybe 5-6 hrs

Figure 5.1: Personal data is heterogeneous because individuals track different information and do so differently. In the MigraineTracker study, participants reported migraine duration in different ways. PT03 reported daily hours while PT08 and PT10 used start and end times under differently named columns. PT09 reported duration as free-form text.

The diversity of the records and their large between and within variations necessitated individualized data processing and visualization pipelines. My collaborators and I developed these pipelines with a combination of traditional programming (Python) and manual visual authoring (Tableau), while frequently consulting participants and incorporating their unique knowledge of their data. Inadequate and limited affordances of commonly available personal data analytics tools drove our decision to use Wizard-of-Oz analysis. This approach served our study but left an important question unexplored: how to support end-users who are not adequately served by the existing tools and do not have the time or expertise to develop goal-aligned pipelines with generic programming and visualization tools?

Existing tools often make assumptions about the structure and semantics of the data rather than accounting for diverse structures and semantics. For example, they assume standard layout, naming, type, or values for the data. They then offer either fixed wrangling, processing, and visualizations [29], or techniques to create new ones given the assumptions (e.g., [146, 183]). However, very limited assumptions can be made about heterogeneous goal-directed data across individuals or for one individual over time as noted above. Closely related to the unwarranted assumptions is the failure of tools in leveraging people's expertise about their own data. Much

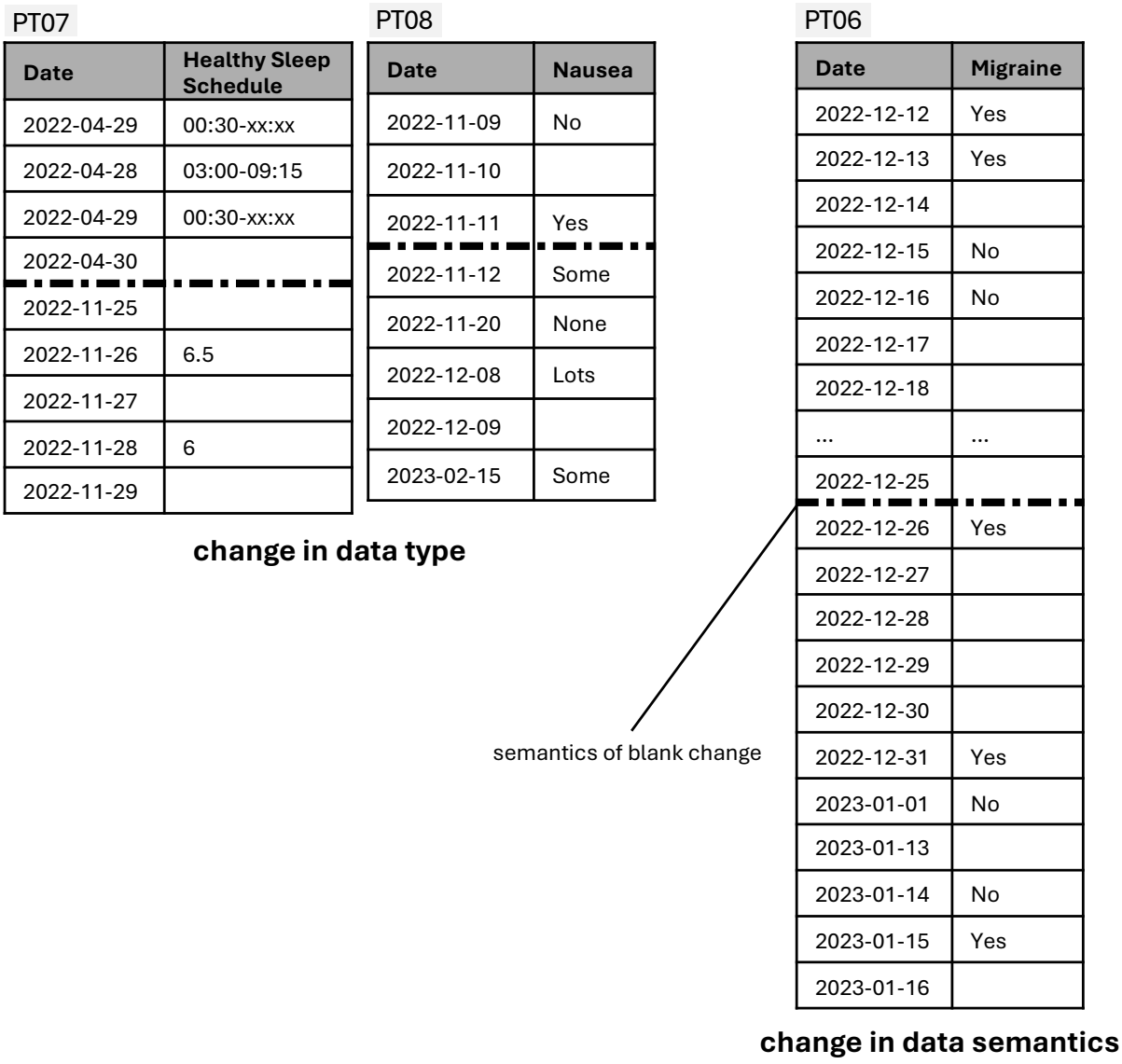


Figure 5.2: Personal data is heterogeneous because goals evolve and people change how they report different information. In MigraineTracker study, PT07 switched from reporting start and end times of sleep to hours of sleep or PT08 switched from 'Yes'/'No' entries of nausea to providing it as ordinal categories. Here the type of data changes, though it appears in the same column. PT06 switched from daily reporting of migraines with 'Yes'/'No', where blanks means missing data, to reporting mostly on migraine days with blanks indicating no migraines. Here the semantics of data changes within the same column.

about the structure and semantics of data can be disambiguated if individuals are given the opportunity to do so. For example, blank entries may indicate missing data for some individuals, whereas absence of the phenomena for others (Figure 5.3). Each individual's knowledge of their data is key to properly handling blanks. This knowledge is sometimes the only reliable way to establish semantics, particularly when the meaning of blank entries changes from missing to absence or vice versa for one individual due to tracking model evolution (PT06 under Figure 5.2).

Past research on end-user data analytics tools offer solutions that can address the challenges posed by heterogeneous data. For example, Wrangler [93] simplifies data wrangling by combining direct manipulation of data with automatic inference, enabling users to iteratively explore and apply operations without prior expertise in data structuring. Data Formulator [192] extends this functionality by integrating programming-by-demonstration or natural language processing, aiming to further reduce the barriers for end users unfamiliar with data science. Unlike traditional analysis tools that require data to be in tidy format, Data Formulator introduces a novel concept binding paradigm that uses an AI agent to automatically perform the necessary data transformations based on author intents, significantly reducing manual data preparation work. While these solutions make it easier for end-users to perform analysis, they assume individuals know what analysis addresses their information needs and their primary challenge is translating the analysis to code. They do not offer support for translating an information need to analysis steps.

Large Language Models (LLMs) allow individuals to enter their information need and get a text or visualization response. As such, they short-circuit the information need-to-analysis translation challenge while handling some aspects of data heterogeneous (e.g., data type variations or data variably named from one individual to another). However, they are often limited in their analysis decisions [81]. Their black-box nature [184] means assumptions they make are typically hidden and there are limited interactions for end-users to check and adjust those assumptions. For example, LLMs are prone to mishandling blank entries, but individuals cannot easily inspect and modify LLM's handling of the blanks.

In this chapter I introduce Analyticons (Analytic + Icons) to enable goal-aligned analysis

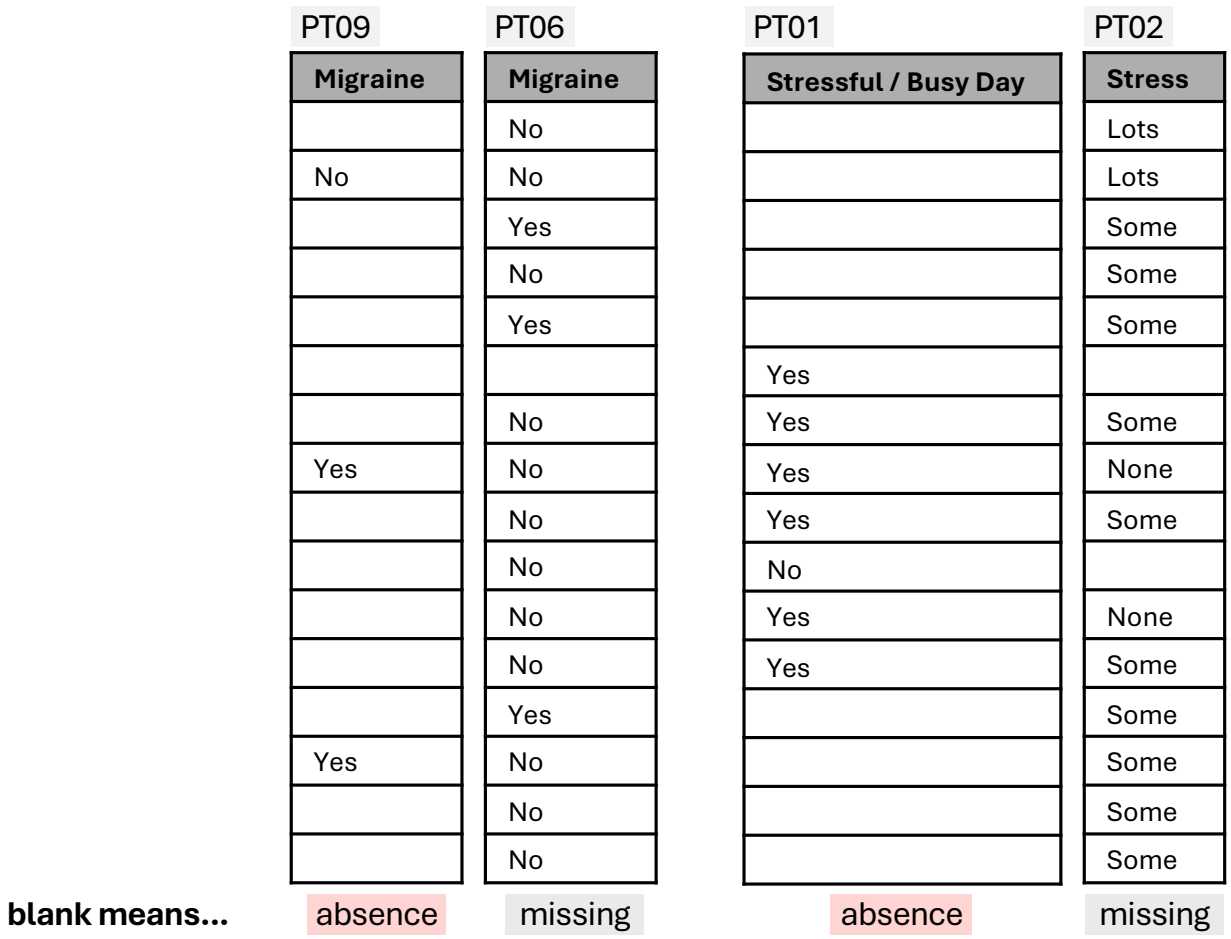


Figure 5.3: Blank entries are challenging to interpret. Some participants use a daily tracking model and explicitly report everyday while others use event based tracking. For the former group, blanks indicate missing data, but for the latter blanks indicate the symptom, treatment, or contributor is absent. For instance, both PT06 and PT09 report migraines. PT06 reports daily. However, PT09 mostly reports the presence of migraines. We observe similar pattern for stress reporting by PT02 and PT01.

of heterogeneous personal health data. Analyticons is an architecture that combines analytics expertise with each individual's expertise about their data. The architecture is based on visual objects of the same name. These objects represent analysis in terms of a set of constraints and adjustments over concepts. As individuals are guided through a sequence of steps to satisfy these constraints, their data is transformed and visualized consistent with their information needs. Informed by specific characteristics and requirements of the domain, Analyticon objects package the analysis expertise beyond what is afforded by tools such as Wrangler [93] or Data Formulator [192] and can thus provide both scaffolding for the analysis steps and their translation into code. People's unique knowledge of their data is factored in the variety of ways they satisfy the constraints. Rather than making assumptions or hiding the specifics of the analysis, Analyticons-based systems offer individuals a range of interactions to control system behaviors and align it to their needs. I share an LLM-enabled implementation of the architecture and demonstrate its expressiveness over a variety of analysis flows. Specifically, I show its ability to support between and within variations in data structure and semantics and its flexibility to handle analysis over new or arbitrary concepts and despite different types of evolution. This is the first of a three-part evaluation planned for an upcoming submission to ACM CHI Conference on Human Factors in Computing Systems (CHI'26). The evaluation not covered in the scope of this thesis involves (1) benchmarking different components of the architecture to characterize their performance and limitations, and (2) a user study to examine how Analyticons support individuals through the end-to-end process of mapping heterogeneous data to diverse information goals.

To summarize, the contributions in this chapter are:

- I characterize the design space of analysis for a set of common information goals in self-tracking of chronic conditions and discuss the utility of this characterization in evaluating the design of personal informatics tools as well as creating translation resources to support designers in making these tools.
- I present Analyticons architecture and constraint-embedded visual objects and describe

an LLM-enabled implementation of it. I also discuss how this approach enables improved support for end-user data analysis and creates new opportunities for it.

- I demonstrate the expressiveness of Analyticons through a representative set of analysis and discuss both ways of expanding the system and its limitations.

I designed and implemented Analyticons based on my re-analysis of MigraineTracker data and process and led the overall research design which was heavily influenced by discussions with Jina Suh, Sean Munson, Jeffrey Heer, and James Fogarty. I thus use the first person plural from here on. In the following sections, I first layout the formative work underlying the design of Analyticons (Section 5.1). I then describe the system with a walkthrough, present key features, and offer the design rationale (Section 5.2). Next, I share a variety of analysis flows Analyticons support (Section 5.3) and discuss implications and promising future directions (Section 5.4).

5.1 Formative Study

We examined patient participant data and the analysis process in their support during MigraineTracker study to better understand and characterize (1) the information needs in using personal data for chronic conditions, (2) considerations and requirements for data analysis to support these needs, and (3) the design space of relevant analytics: the key constructs, tasks, and operations for data preparation and visualizations. This examination identifies a design space for end-user analysis of personal data and points to design goals within this space. Subsections below detail findings along these lines and design goals informing Analyticons.

5.1.1 Information Goals in Managing Migraines

We analyzed patient information goals (i.e., questions they wanted to answer with tracking data) in migraine self-tracking with the ultimate goal of better understanding information needs in chronic condition management. Our analysis finds the majority of information goals fall under four major categories: (1) how do symptoms vary? (2) how do preventive treatments work? (3) how do abortive treatments work? (4) how are symptoms related to a factor? These findings are

consistent with Schroeder et al. [172]'s work and extend it. They reported nine types of questions in the context of self-experimentation that can be broadly considered under the fourth category of information goals above. This set of four information goals is thus a reasonable starting point for exploring the space of analytics constructs relevant to chronic condition management. However, we acknowledge that the four categories we found are not comprehensive. We expect additional categories with further explorations into the topic of information needs in chronic condition management and with a careful examination of conditions with different characteristics compared to those we and Schroeder et al. studied.

As we present the four types of information goals below, we highlight analytics constructs relevant to them (italicized). These constructs, along with the information goals, informed the design of Analyticons and include: types of variations, units of time, windowing for comparisons, change of reference, thresholding, conditioning, and a variety of temporal alignments (e.g., same-day vs. within a few days, lagged vs. cumulative vs. let-down). Figure 5.4 Summarizes the information goals and their analytics constructs.

How Do Symptoms Vary?

Patients wanted to better understand variations in the presence (i.e., frequency), levels (e.g., length or strength), or values (e.g., location or type of pain) of their symptoms. Common examples include 'how often do I get the symptom (e.g., migraine)?', 'how long do symptoms (e.g., headaches) last?', or 'where (i.e., head locations) do I get the symptoms?'. Patients were interested in both the overall variations (e.g., total days of the presence of migraine vs. its absence or the distribution of migraine durations) and the variations over time. Depending on the specifics of the question, different units of time were desired (e.g., frequency of migraine symptoms over days of the week, weeks of the months, months of the year, or seasons). *Types of variations* (i.e., presence, levels, or values) and the different *units of time* are the key constructs to note here. Patients sometimes specified additional *conditions* for the variations of interest (e.g., 'how bad morning migraines are typically?'), another construct we observed in this and other categories of information goals.

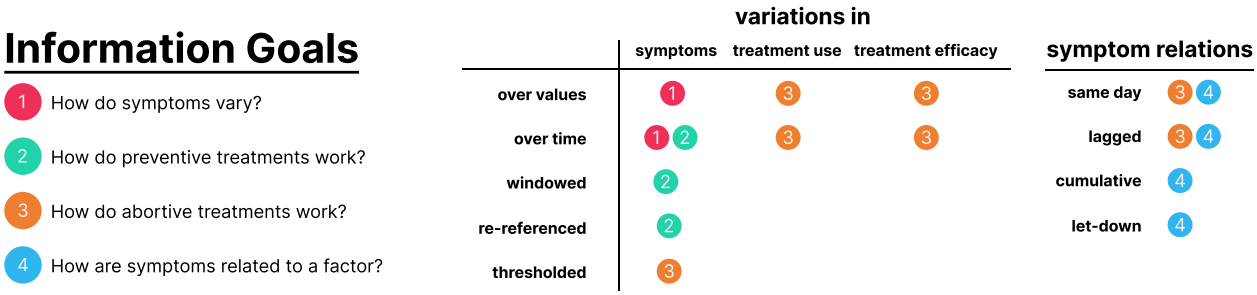


Figure 5.4: Information goals and their related constructs. Goals are listed on the left. Their related constructs appear in two tables on the right with goals marked for the constructs with which they were associated. In summary, variations in symptoms and (abortive) treatment use or efficacy were desired over the values of the symptom or treatment variable as well as over time for different units of time. Sometimes, comparisons were desired over specified windows of time (e.g., before or after a new treatment). Variations were sometimes in reference to specific days (e.g., days of preventive treatment). Thresholding symptom values was sometimes desired to define presence or absence of a certain state (e.g., recovery). The relation between symptoms and other factors could be described with same-day, lagged, cumulative, and let-down occurrence. Lagged occurrence describes presence of one variable within a certain number of days since the presence of the other variable. Cumulative is the occurrence of one variable after multiple days of the other variable's presence. Let-down concerns with the occurrence of one variable when multiple days of the presence of the other is followed by its absence.

How Do Preventive Treatments Work?

Patients were generally interested in understanding whether preventive measures they were regularly taking affected the presence, levels, or values of symptoms over time. The most common case under this category was to consider symptom variations before or after starting or changing treatments (e.g., starting a new medication, changing a medication's dosage). However, *windowing* time based on a treatment was more broadly of interest even when the windows were disjoint (e.g., windows of regular chiropractic therapy) as patients were hoping to make comparisons within and outside of the defined windows (e.g., 'how symptoms change with chiropractic therapy twice a month?'). In addition to the windowing construct, patients were interested in considering symptom variations *in reference* to the treatment days. This was most common for medications patients received once a month or once every three months, as in 'how are symptoms within a month since every Ajovy [a type of treatment] shot?'

How Do Abortive Treatments Work?

Understanding usage characteristics and efficacy of treatments used as needed to abate symptoms was of high importance for patients. They often wanted to characterize treatment use in terms of usage (i.e., taken vs. not), amount (e.g., dosage), and timing (e.g., on early signs of symptoms vs. not) overall and over time. When treatment efficacy was directly reported within the app, patients were interested in the frequency and levels of efficacy both overall and over time (e.g., 'how often does the medication work?'). Alternatively, *same-day* presence or levels of symptoms after abortive treatment was desired overall or over time (e.g., 'how long do symptoms last after treatment?'). Sometimes, patients wanted information for a given *thresholding* as an indication of treatment efficacy (e.g., 'how frequently symptoms last for over an hour?'). They were also interested in comparing symptom variations (i.e., presence, levels, or values) between treated and untreated days. Other comparisons of interest were in terms of presence, levels, or values of symptoms with respect to different amounts and timings of treatments both on the same day of the treatment and *within a few days* of the treatment (e.g., 'are there bounce-back symptoms a day after treatment?'). Patients often used different treatments and were interested in usage and efficacy information for each treatment individually and for all treatments combined.

How Are Symptoms Related to a Factor?

Patients were interested in the relationship between the presence, levels, or values of symptoms given the presence or levels of a given factor (e.g., 'How likely is it to get symptoms on days when I am dehydrated?'). In addition to concurrent (e.g., same-day) relations, patients considered *lagged*, *cumulative*, or *let-down* relations. Lagged relations considered symptoms a certain number of days since the presence of a factor (e.g., 'do I get symptoms a day after drinking alcohol?'). Cumulative relations were concerned with symptoms after multiple days of a factor presence (e.g., 'do I get symptoms after two or more days of poor sleep?'). Let-down relations asked about symptoms when consecutive days of factor presence was followed by its absence (e.g., 'do I get symptoms when three or more stressful days are followed by a relaxed

day?'). Relations of interest sometimes involved more than a single factor (e.g., 'do I feel more fatigued on stressful days and do I get more symptoms when fatigued?'). When uncertain of the relevant factors, patients wanted to discover them (e.g., 'what combination of contributors co-occur?', 'how do different contributors add up?').

5.1.2 Domain-Specific Considerations and Requirements

For the ultimate goal of supporting end-user analysis of personal health data, it is useful to examine what distinguishes this type of analysis from others. High between and within variability in both data and information needs is one key differentiating factor and requires personal data analytics tools to accommodate the variability. Another distinctive factor is each individual's unique knowledge of their data, which can be utilized as a source of expertise in the analysis. We complement the earlier notes on these aspects with further details below.

Data and Needs Widely Vary

We described the heterogeneous nature of participant data in the opening of this chapter: patients recorded different information, recorded the same information differently, and one person changed their recording over time due to goal evolution. In addition to individualized data collection, each patient considered a unique set of information goals; not all goal types were relevant to a single individual and there were variations for a given information goal (e.g., one person was interested in the presence of a given symptom while another was interested in its levels. Another person was initially interested in presence and then became interested in levels).

Individuals Are Experts of Their Own Data

Individual expertise is often the key to managing the variability and ambiguity associated with the structure and semantics of personal data. Below are three representative cases of the critical role an individual's knowledge play in guiding data processing. Case of handling blank entries was noted earlier and is repeated for a consolidated presentation.

Handling Blank Entries. As we noted earlier, getting the semantics of blank entries is challenging. For example, one person reported stress levels everyday. Another person reported it when they were stressed. There is no single solution for handling blanks; it means missing data for the first person and no stress for the second. This knowledge, one way or another, rests with the user.

Designating Data Sources. One person reported non-migraine headaches directly by responding 'Yes/No' under the named item 'Non-Migraine Headache'. Another person implicitly reported it by responding 'No' to whether migraine happened but providing a rating for migraine severity. In the latter case, user's knowledge is critical in knowing where to find relevant information and how different pieces of information are related to each other and to various concepts of interest.

Handling Outliers. Non-migraine headache rating of 6 is an outlier for a person who considers headaches above 5 a migraine. The same value is considered normal for another person who considers headaches 8 and above or the presence of other symptoms such as light and sound sensitivity as migraine. User knowledge is necessary for finding and handling outliers.

5.1.3 Operations

Examining the analysis steps in Wizard-of-Ozing data preparation and analysis, we identified a set of analysis tasks along with operations to perform them. Wrangling and transformation tasks included identification and cleanup of inconsistencies, errors, outliers, or missing data, recoding of values, extraction of new values, filtering, slicing and segmentation, grouping, and aggregation. Data interaction tasks were search, annotation, and alignment of different data streams. Operations to perform these tasks are summarized in Table 5.1.

5.1.4 Design Goals

We laid out the space of information goals and their associated analytics constructs. We also highlighted common analysis tasks and their underlying operations. The design objective is to enable the expression, sequencing, and specification (e.g., parameter setting) of these operations consistent with information goals while meeting the specific requirements of our setting (i.e., high

Table 5.1: Operations performed in data analysis during MigraineTracker study. These are well-defined operations and are found in other analysis tools (e.g., SQL databases) primarily intended for computer scientists and data analysts, rather than end-users.

Operation	Input Signature	Description
find	(records, rule → {category}n, [{column}], [{value}])	Assigns each record to a category in a set of N categories based on the criteria defined by the rule function applied to a set of columns. It optionally returns records whose categories are in the set of values, i.e., it returns a subset of records.
remove	(records, [{column}])	Removes records of columns.
replace	(records, rule → {value}, {column})	Replaces contents of each record under column with a value based on the mapping function rule.
split	(records, rule → [{value}], {column})	Returns for each record under column a set of values according to rule.
combine	(records, rule → {value}, [{column}])	Returns a value for each record based on the rule over a set of columns.
aggregate	(records, rule → value, {column})	Returns a value based on the aggregation rule for records under column.
reindex	(records, rule, column)	Changes the index of records under column based on the rule.
review	(records, {column})	Presents records under columns.
assign	(record, {column}, {value})	Assigns a value to a column of a record.
select	(records, [{column}])	Select records under one or more columns; if either record or column are not given (e.g., select([{column}]) or select(records)) all values are returned.

variability) and leveraging its distinct features (i.e., individual expertise). This can be summarized as the following design goals.

DG1 *Enable End-to-End Analysis.* Provide scaffolding for individuals engaged in data analysis around *what* steps to take and *how* to take them. This scaffolding leverages data content and stated information goals to map analytical intents to steps that once taken address the information goal. Scaffolding should also include alerting individuals of the potential mismatch between data and information goals (i.e., breakdowns in steps from where the data is to where it should be).

DG2 *Integrate Different Expertise.* Combine analytics expertise (packaged and presented to individuals as different forms of scaffolding) with user-expertise where individuals give input and can review and repair system decisions (e.g., inference or recommendations).

DG3 *Accommodate High Variability.* Assume minimal standards for data (both structure and semantics), a wide range of goals, relaxed expressions of those goals, and varied preferences for data representations (e.g., visualizations).

DG4 *Build Goal-Aligned Support.* Create support for the key constructs underlying four categories of information goals to bridge the gap between computationally well-defined operations and the information needs. Such support should essentially package a sequence of operations for common analytics tasks.

5.2 Analyticons Architecture

We designed Analyticons as an architecture for end-user goal-aligned analysis of heterogeneous personal health data, where analytics expertise is packaged to support information goals, and where individual's intimate knowledge of their data is integrated in applying this expertise to transform and visualize the data. Below, we first illustrate how an individual experiences Analyticons by walking through a representative example of data analysis within a system implementing this architecture (Section 5.2.1). Following an overview of the key components of the system (Section 5.2.2), we describe the Analyticon objects (Section 5.2.2), which are at the core of the architecture. Design rationale for the architecture, the underlying objects, and system components are offered throughout the description.

5.2.1 Analyticons at a Glance

We use PT10's data from MigraineTracker study for the commonly asked question 'how long do my headaches take?' to illustrate Analyticons's interaction flow (Figure 5.5). Having provided the data and information goal (i.e., the question), individuals receive recommendations for types of visualizations that can help address the question (Figure 5.6), followed by detailed instructions

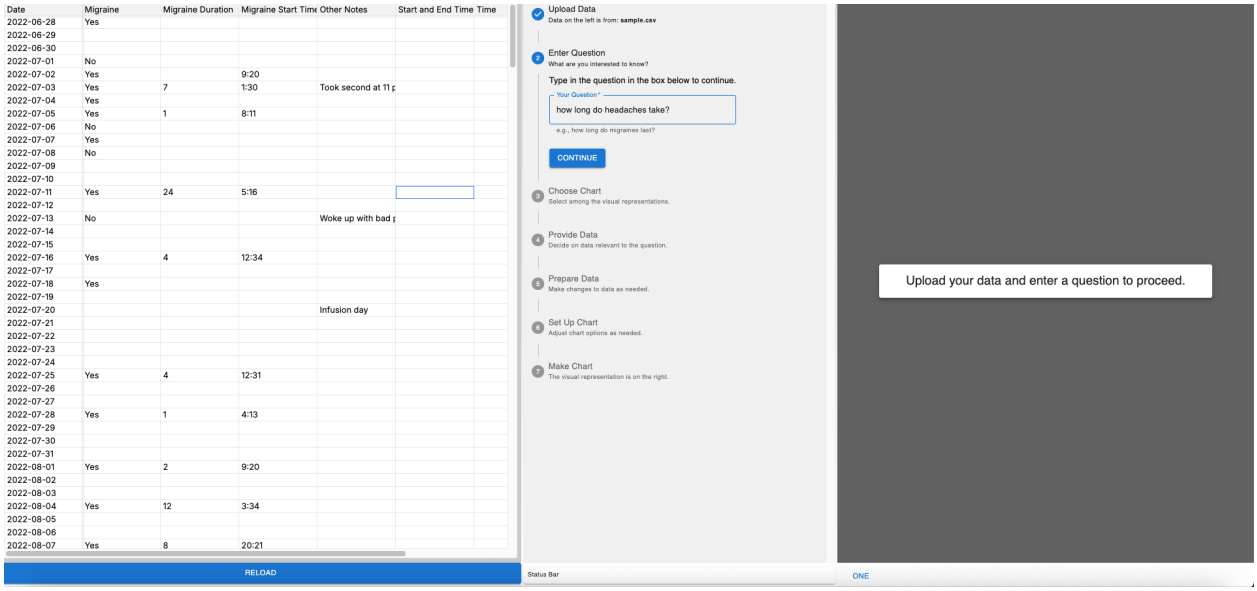


Figure 5.5: Individuals are prompted to enter an information goal (i.e., a question) they want to answer with their data.

that guide them in generating the visualization with the data (Figure 5.7, b). Data is visible at all times in tabular format with rows representing calendar dates and columns representing different symptoms, treatments, and contributors (Figure 5.7, c). The desired visualization appears on the right upon selection, though it is labeled as a 'work in progress' to indicate it is not yet reflecting individual's data (Figure 5.7, a).

Concept Definition. Assuming a bar chart of duration is of interest, the user is first asked to confirm or revise system's inference of what the question is about: the *concepts* underlying the question (e.g., 'duration' in Figure 5.8). The rest of the interactions with the system and its computations revolve around the concepts; concepts become the shared units of operation between the individual and the system.

Concept Mapping. With concepts defined, the next step is confirming or revising the columns that the system thinks contain information relevant to the concepts (Figure 5.9). In our example, the system has identified the column 'Migraine Duration' in relation to the 'duration'

Figure 5.6: Individuals receive recommendations for visualizations that can address their information goals.

concept (Figure 5.9, a). Drawing up on her knowledge, the individual adds 'Start and End Times', which she has also used to record duration related information (Figure 5.9, b). The system updates column highlighting to provide feedback of the shared understanding to the individual. Note the individuals can also add new columns to their data and enter values if they desire. This is most helpful when individuals have information in some other source (e.g., their calendar) they would like to bring to their analysis.

Concept Valuation. Having identified the concepts and where to find relevant information about them, individuals are asked to specify concept values (Figure 5.10). They can provide multiple

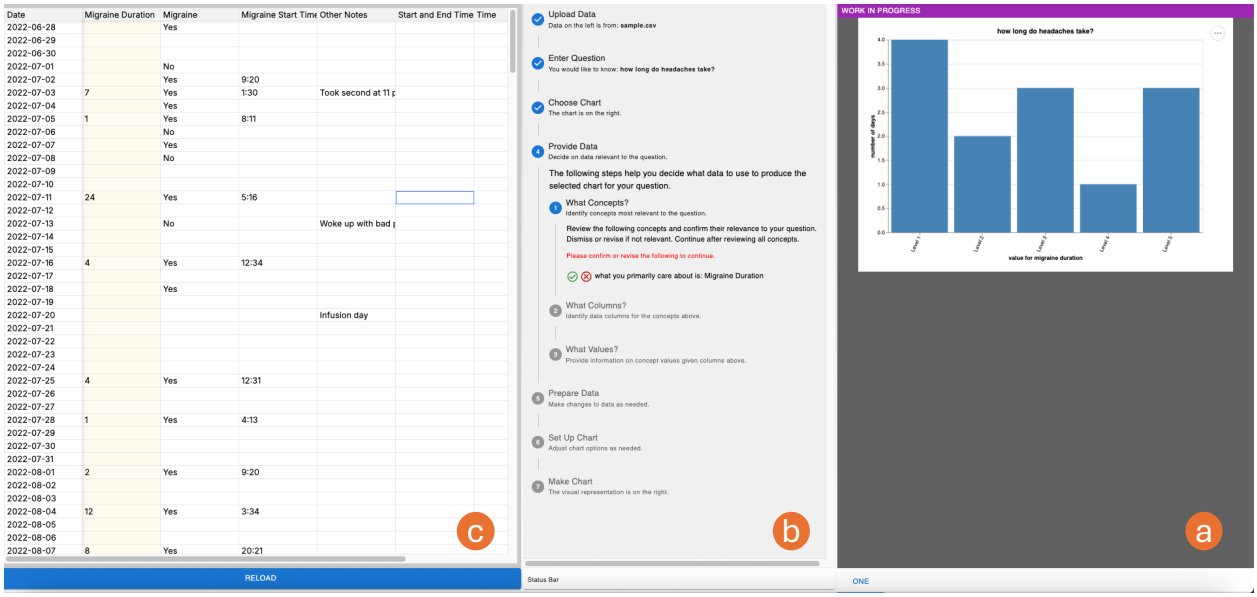


Figure 5.7: Individuals are guided through a set steps to prepare their visualization of choice based on their data. (a) visualization to reflect individual's data as the data is transformed through the process appears on the right. The top banner indicates it is still being constructed and is a work-in-progress. (b) detailed instructions that help individuals use their expertise to transform the data appear at the center. (c) data assumed to be tabular appears on the left. Its rows correspond to calendar dates and columns contain data for symptoms, treatments, and contributors on each day. System highlights parts of the data it thinks are relevant to the information goal ('Migraine Duration' column).

specifications of a single concept, each applicable to a segment of the data, which is a period of time where the specification is valid. In our example, the individual leverages her knowledge of the change in recording duration information in the middle of their tracking on '2022-02-16', when she switched from recording length of migraines under 'Migraine Duration' to recording the start and end of migraines under 'Start and End Times' (Figure 5.10, a). She has the ability to adjust the relevant columns for each segment/period (Figure 5.10, a) and specify duration values in each: she opts to use duration values as-is prior to the change on '2022-02-16' by copying them all from 'Migraine Duration' to the column now dedicated to duration concept (Figure 5.10, b). For the other period (i.e., after '2022-02-16'), she specifies values by giving examples of how her information under 'Start and End Times' map to duration values and asks the system to fill the

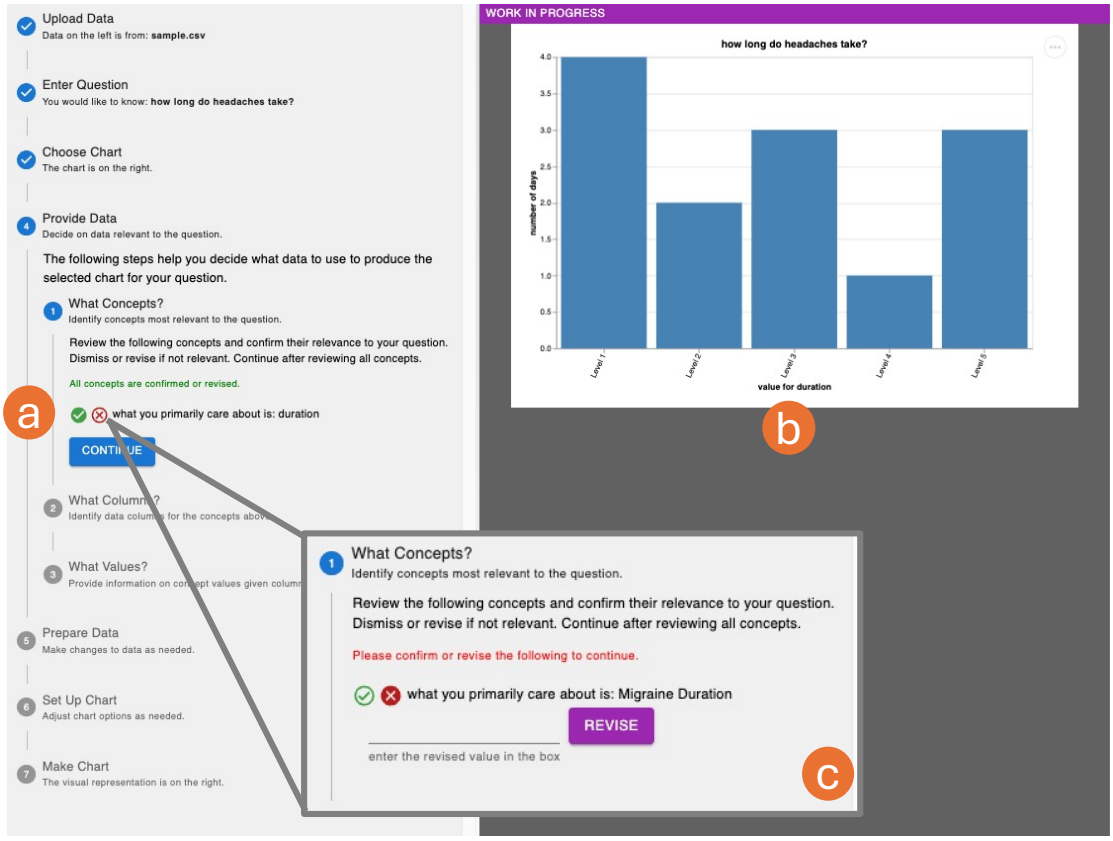


Figure 5.8: Individuals are asked to confirm or revise the concepts system has identified in their question. (a) concept of 'duration' for the question 'how long do headaches take?' is confirmed by the individual indicated by the fully green check-mark. (b) the concept is also reflected on the visualization within the title of the y-axis. (c) individuals have the option of revising the concepts system has identified to better reflect their need.

rest (Figure 5.10, c-d). She can review system output and approve or reject them, then have the system try again (Figure 5.10-e). User demonstrations and system-generated values are colored and formatted accordingly. This step supports different aspects of data heterogeneity based on each individual's needs and understanding of their data as we further illustrate in Section 5.3.

Refinements and Adjustments. After concept values are assigned, the system checks if they need any additional processing before using them in the visualization. This is done by examining visualization-specific constraints (Figure 5.11, a) and providing individuals with means and

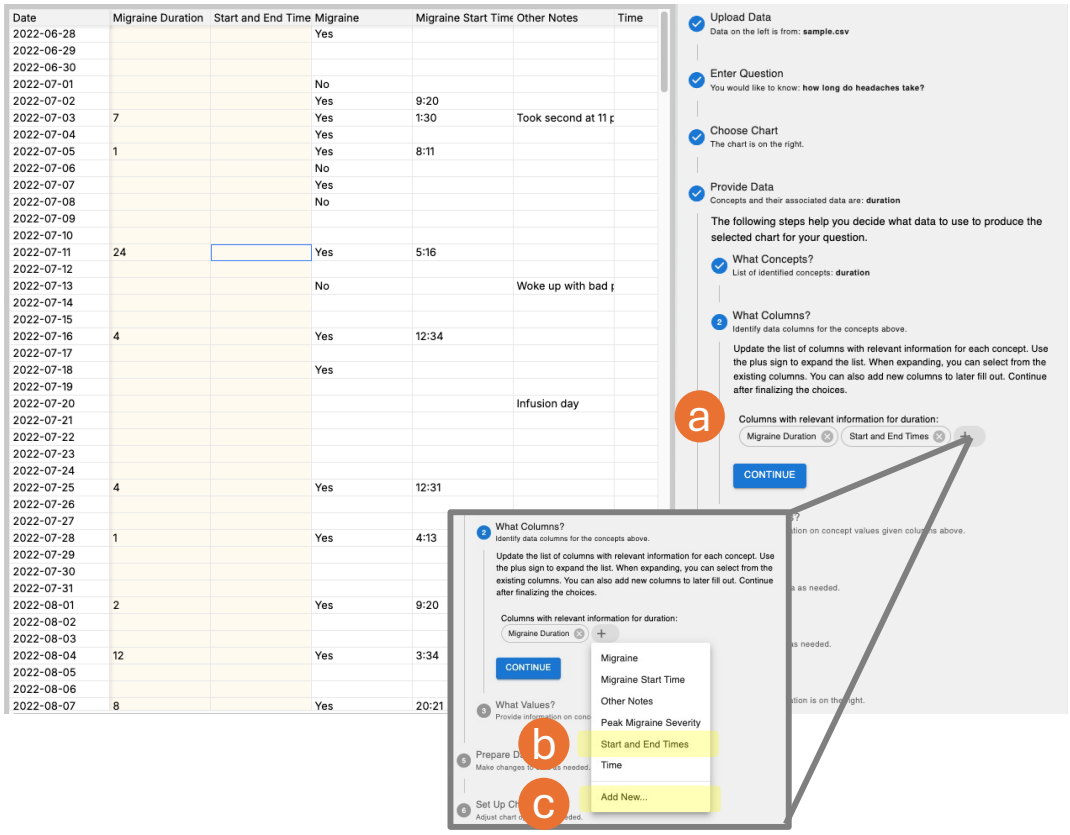


Figure 5.9: Individuals can confirm or revise data columns the system has identified as relevant to each concept. (a) system indicates 'Migraine Duration' to contain information relevant to 'duration' concept. (b) the individual adds 'Start and End Times' as another column they know has the relevant information. (c) there is the option of adding new columns to data.

instructions to address them (Figure 5.11, b). This can again happen at the scale of periods when needed. Here, the same slider widget is relevant in both periods. Visualizations provide a representation of the data in relation to the question after the transformations and adjustments are in place (Figure 5.12). In our example, the visualization expects distinct categorical amounts of duration and offers a slider widget for specifying the categories, thus transforming numeric values to ordinal levels suitable for generating the visualization. Individuals have full control over what ordinal levels are relevant to them.

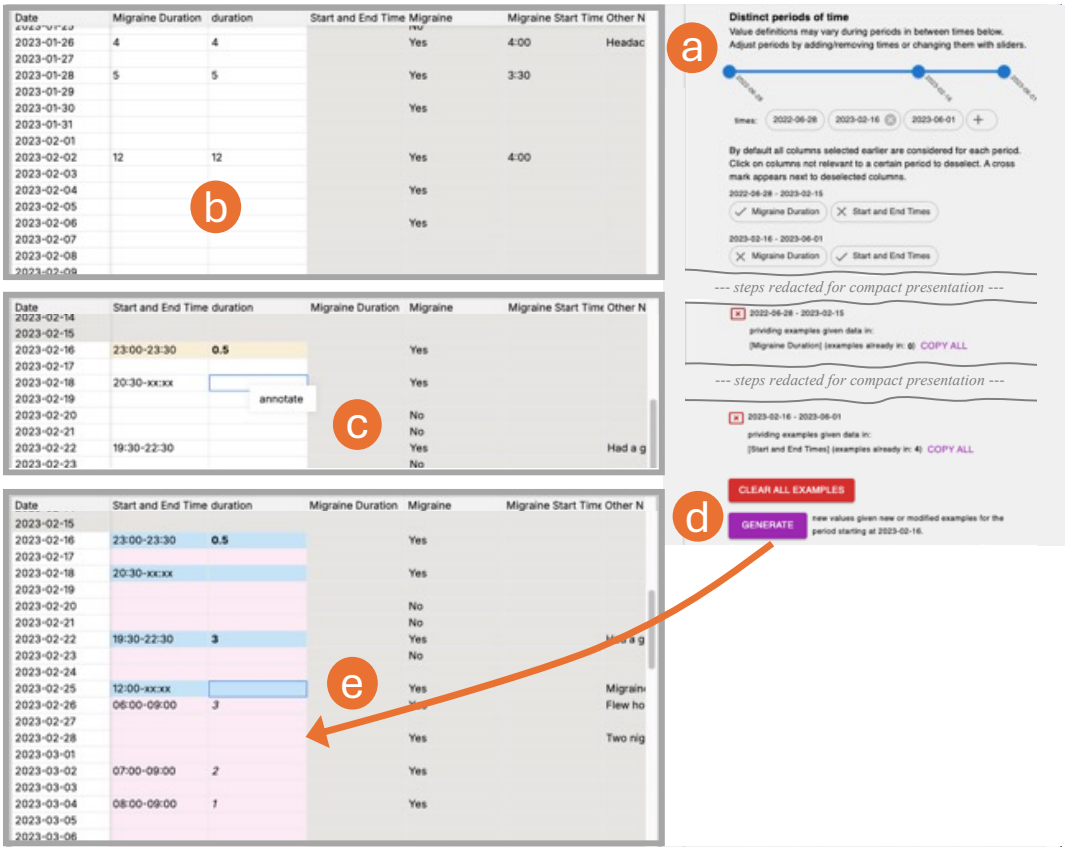


Figure 5.10: Individuals are asked to specify concept values given the relevant data. (a) they can provide different specifications by specifying times each specification is applicable to. Different columns may apply to each period of time. (b) individuals can use values as is by copying them from the relevant columns, or (c) can extract new values from those columns. This is achieved through providing examples and (d) having the system use the examples to generate remaining concept values as in (e).

5.2.2 System Components

The system consists of four key components (Figure 5.13) to help individuals find answers to a given question using their data: analysis selection through mapping to Analyticons, source and value satisfaction, and parametrized adjustments. Three of the four components shadow the main constructs of Analyticon objects that are at the core of the architecture. Data is assumed to be in tabular form with rows representing calendar dates and columns associated with reports of symptoms, treatments, and contributors on each day. The system is implemented as a desktop

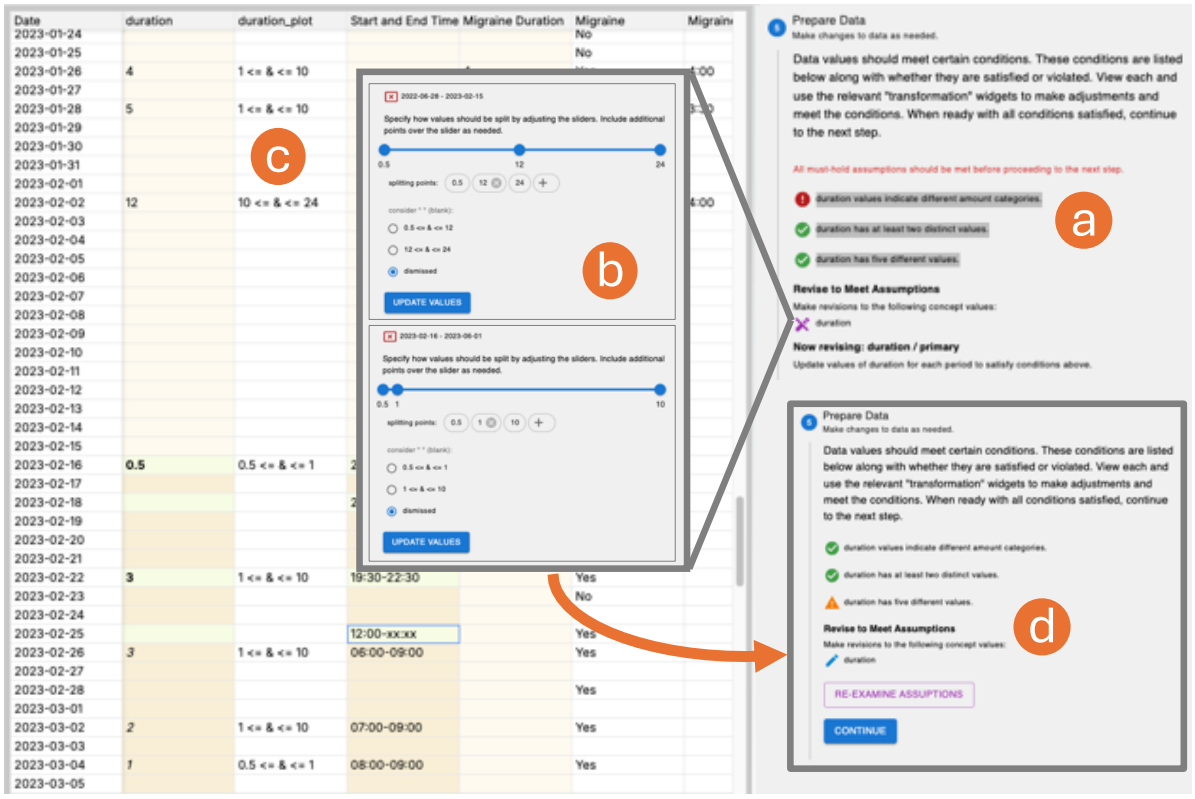


Figure 5.11: Individuals need to ensure concept values meet certain requirements. (a) they are informed of these requirements and whether they are met, and (b) are offered means of addressing the requirements as in (c). They can proceed to generating their visualizations when must-meet requirements are met.

web application in React with a Flask backend. Visualizations are built with Vega. Below, we further describe Analyticons objects and system components.

Analyticons: Goal-Aligned Constraint-Embedded Visual Objects

Analyticons are closely tied to information goals (**DG4**). Each Analyticon corresponds to an instance, or subtype, of an information goal. Instances are differentiated by the relevant analytics constructs. For example, there are instances of 'how do symptoms vary' for different types of values (i.e., nominal, binary, ordinal, or numeric) and whether variations are considered over time (Figure 5.4). An Analyticon has three primary constructs: source constraints, value

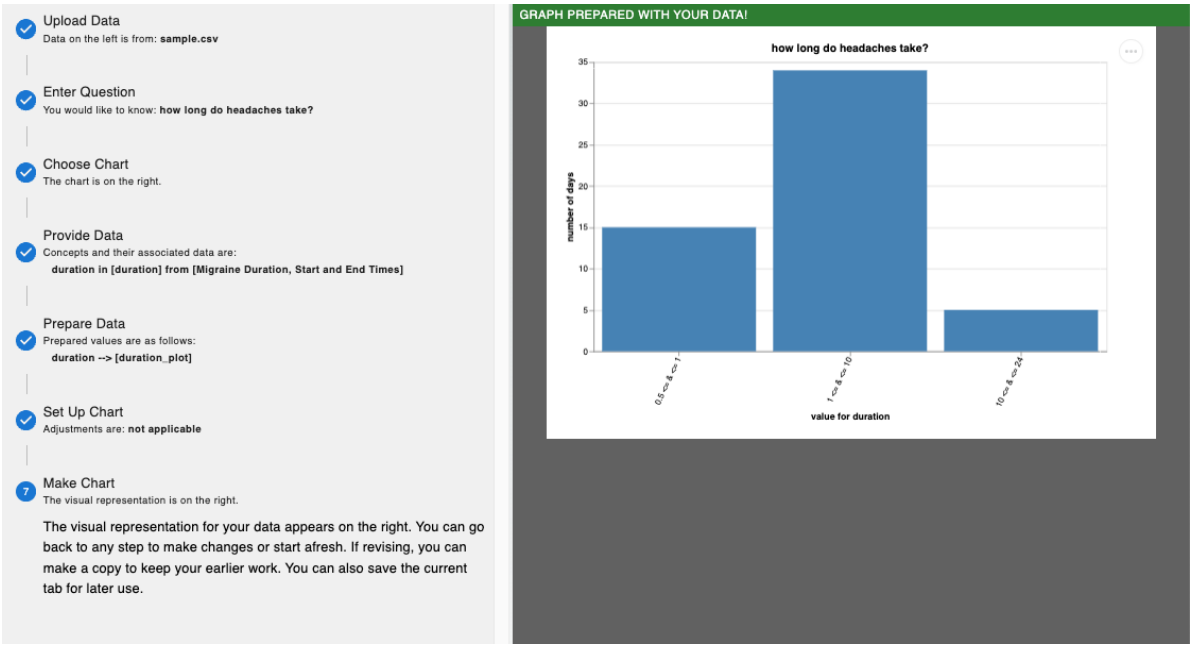
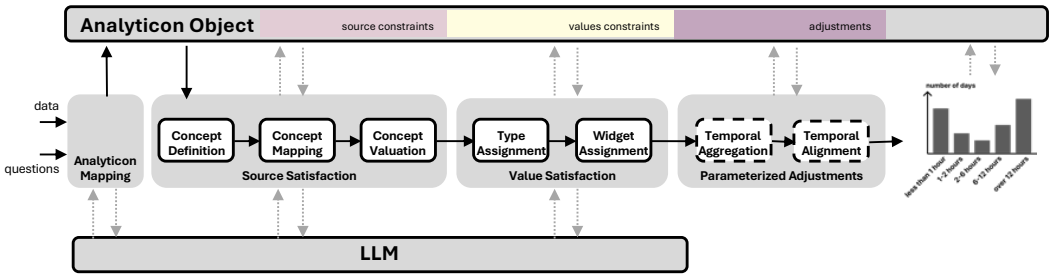


Figure 5.12: Visualization prepared for 'how long do headaches take?' of the example in Section 5.2.1. The green banner shows the visualization represents individual's data.

constraints, and adjustments. Figure 5.13 shows the Analyticon for the example we walked through in Section 5.2.1, annotated with its constructs. These constructs are defined at the level of concepts rather than data which allows a wide coverage despite within and between variations in the records (**DG3**). Concept level constructs also enable flexible adaptations by individuals given their knowledge of the data (**DG2**).

Source constraints get at *what data is relevant*. In our example, there is a single variable of interest: duration (of migraines). Additional variables matter for other types of information goals and their associated Analyticons will reflect that difference. Value constraints are about *how the data should look like*. In the example, concept values are expected to be ordinal categories. Parameterized adjustments are *domain-specific special operations*. Not all Analyticons include these operations. The Analyticon in our example has none. Aggregation by time is the most basic of the special operations given the timeseries nature of data. Different types of temporal alignment between variables apply given analytics constructs of this particular domain




```
{
  "IGtype": { "tag": "IG1", "txt": "how does a symptom variable vary?" },
  "IGsubtype": { "tag": "SMP-VAR-TOT-O", "txt": "variations in levels of ordinal categories"},
  "assumptions": {
    "source": [
      {
        "tag": "primary",
        "explanation": "the concept we primarily care about",
        "info": {"status": "UNMET"},
      }
    ],                                                          source constraints
    "authoring": [
      {
        "tag": "ordinal",
        "explanation": "source column contains non-numeric ordinal values",
        "info": {"cost": "INF", "on": "primary", "status": "UNMET"}
      },
      {
        "tag": "distinct",
        "explanation": "source column has at least two values",
        "info": {"cost": "INF", "on": "primary", "status": "UNMET"}
      },
      {
        "tag": "levelNum5",
        "explanation": "source column has five levels",
        "info": {"cost": "100", "on": "primary", "status": "UNMET"}
      }
    ],                                                          value constraints
  }
  "adjustment": [],                                             adjustments
}
```

Figure 5.13: Components of Analyticon system and the Analyticon object of the example in Section 5.2.1. There is a component for suggesting an Analyticon given data and information goal (i.e., the question), along with three other components that match Analyticon object constructs: source satisfaction → source constraints, value satisfaction → value constraints, and parameterized adjustment → adjustments. The system guides individuals in meeting the constraints and making the adjustments. Some of the system functionality is powered by LLMs.

and include same-day, lagged, cumulative, and let-down. Put together, each Analyticon is a combination of different source and value constraints along with adjustments. As individuals go through the steps for meeting the constraints and specifying the adjustment parameters, their data is transformed. These transformations take the data from where it is to where it should be for the visualization addressing the information goal (**DG1**).

Each Analyticon packages a goal-aligned analysis and makes it available to individuals to adapt (**DG4**). This packaging draws on specifics of the data and domain to offer end-users

concept-level rather than *computational* operations typical to computer or data scientist-oriented tools (i.e., the kind of operations in Table 5.1). The other components of the system provide additional scaffolding around the analysis steps packaged by an Analyticon (**DG1**): *what-to-do* scaffolding of **DG1** is achieved via Analyticon objects and the system guidance from source satisfaction to parameterized adjustments and *how-to-do* scaffolding of **DG1** is achieved through the interactions in each of these steps (e.g., via approvals/revisions, transformation widgets).

Mapping Questions to Analyticons

A given information goal along with the individual's data identifies a subspace of analytical intents. This subspace is mapped to a set of Analyticons (**DG2** - analytics expertise). We use large language models (LLMs) to match the input (i.e., individual's question and data content) to information goals and their subtypes (Figure 5.14). Any information goal relevant to the input is offered among recommendations for the individual to choose from (**DG2** - user expertise). To allow maximum flexibility (**DG3**), individuals can browse the full *gallery* of Analyticons for selection (Figure 5.6). The top-down approach of starting with a visualization has been commonly reported in human-centric visual authoring tools (e.g., [116]). Anecdotally, MigraineTracker study patients demonstrated a similar top-down tendency. Especially when refining their information goals, they requested visualizations resembling the ones they had recently seen with new concepts or modifications to the same concepts.

Satisfying Source Constraints

With an Analyticon selected, the system guides individuals through satisfying constraints (**DG1**) starting with source constraints. Because Analyticons are structured around concepts, the first step is *grounding the concept.* This is achieved through three subcomponents for defining concepts, mapping them to data, and specifying concept values given the data (illustrated in Section 5.2.1 and Figures 5.8, 5.9, 5.10). These components are driven by Analyticon information (Figure 5.13) and the individual's input (**DG2**). Each subcomponent does one or both of the following: making inferences and asking individuals to verify and revise as needed (e.g., in concept definition and

```
You are a data analysis assistant. You are given the following data and are asked to decide on
the type of analysis given a question. Consider the data content in making the decision. For
example, when asked about the frequency of something happening, any report of it may count
(binary, ordinal, numeric). But when asked about certain changes, binary values may not be
enough. Provide a yes/no response.

data:
{data}

examples:
the question is "how long do headaches last?". Is this question about "how does a symptom
variable vary?" with "variations in levels of ordinal categories"? : yes,
the question is "how long do headaches last?". Is this question about "how does a symptom
variable vary?" with "variations in nominal values"? : no,
the question is "how long do headaches last?". Is this question about "how does a symptom
variable vary?" with "variations in presence over time"? : no,
the question is "how long do headaches last?". Is this question about "how does a symptom
variable vary?" with "variations in levels of ordinal categories  over time"? : yes,
the question is "how long do headaches last?". Is this question about "how does a symptom
variable vary?" with "variations in numeric levels over time"? : yes,
the question is "where do headaches hurt?". Is this question about "how does a symptom variable
vary?" with "variations in levels of ordinal categories"? : no,
the question is "where do headaches hurt?". Is this question about "how does a symptom variable
vary?" with "variations in nominal values"? : yes,
the question is "how often do I have migraines?". Is this question about "how does a symptom
variable vary?" with "variations in presence over time"? : yes,
the question is "how long do headaches last?". Is this question about "how is a symptom variable
affected by a factor?" with "same day presence of symptom and factor"?: no

the question is "{question}". Is this question about "{IGtype}" with "{IGsubtype}"?
```

Figure 5.14: LLM prompts for assigning Analyticons to questions. The text in blue would be filled programmatically. 'IGtype' has the information about the type of the information goal the Analyticon object addresses. 'IGsubtype' specified the instance of the information goal the Analyticon supports.

mapping, or over values generated given examples), or providing detailed instructions to guide the individual's choices (e.g., in specifying periods within concept valuation). Given the emphasis of Analyticons on concepts, interactions between individuals and system are around concepts. That is, concepts become the shared units of operation between people and the system, agnostic to the underlying data (**DG3**). System inference is driven by LLMs that use information in each Analyticon along with the data and question to generate answers (Figures 5.15, 5.16, and 5.17).

Satisfying Value Constraints

After relevant concepts are grounded, the system checks if their values meet constraints set by the Analyticon (**DG2** - analytics expertise) and asks the individual to further refine concept values

```
You are a data analysis assistant. You are asked to help the user find answers to questions like
"{IGtype}" where {IGsubtype} are considered. You are given the question and asked to complete a
task on it given a data table with the following columns is used to answer the question: {data
columns}. Complete the task in less than 5 words.

Example task and output:
Task: find the symptom concept that the question below asks about.
Question: do I get earlier migraines with poor sleep?: timing

Task: find the symptom concept that the question below asks about.
Question: {question}
```

Figure 5.15: LLM prompts for extracting concepts underlying a question (i.e., an information goal). The text in blue would be filled programmatically. 'IGtype' has the information about the type of the information goal the Analyticon object addresses. 'IGsubtype' specified the instance of the information goal the Analyticon supports.

if needed (**DG2** - user expertise). In the example of 5.2.1, these constraints require values to be ordinal with at least two distinct levels (Figure 5.11, a). These are must-meet constraints and are identified by red circle warning sign given the cost associated with violating them is infinite (Figure 5.13). The other constraint of the example (having five different levels), identified by orange triangle warning sign, has a finite cost and the final visualization is valid even when this constraint is not met. The system also informs individuals when a certain constraint cannot be met (e.g., the data is binary whereas ordinal values are expected) and requests the user to consider revising their concept grounding (**DG1**). Constraints related to types of values depend on determining the type of values, which uses LLMs when deciding if values are nominal or ordinal categories given the flexibility of these types. All other data types (i.e., binary, numeric, time, time range, and text) are assigned via traditional programming.

When constraints are violated, the system automatically identifies the types of transformations needed, assigns the interactive widget that can help the user define the transformation, and initializes the widget as needed (**DG1** and **DG2**). In the example, the system provides a slider to help the individual with numeric to ordinal transformation (Figure 5.11, b). Transformations should be specified in each period because value types may vary in different periods of time. This is most relevant when patient recording evolves within one column and they record values of different types in different periods of time (**DG3**).

```
You are a data analysis assistant. You are asked to help the user find answers to questions like
"{IGtype}" where {IGsubtype} are considered based on a data table with the following columns:
{data columns}. You will be given user's question and a concept of interest in the question. You
will consider the data columns and their content to suggest the columns to use for the concept.
The answer is usually one column but sometimes multiple columns should be used. If the answer
multiple columns, separate them by commas. No duplicates are allowed. Make sure each column
exists in the data table.

Example:
Given data for columns Migraine, Peak Migraine Severity, Nausea, Migraine Start Time, Light
Sensitivity, Migraine End Time
data:
   Migraine: [Yes, No]
   Peak Migraine Severity: [2, nan]
   Nausea: [Yes, No]
   Migraine Start Time: [12:00, nan]
   Light Sensitivity: [Yes, No]
   Migraine End Time: [18:00, nan]
What columns to use for timing?: Migraine Start Time, Migraine End Time
Given data for columns Migraine, Peak Migraine Severity, Nausea, Migraine Start Time, Light
Sensitivity, Migraine End Time
data:
   Migraine: [Yes, No]
   Peak Migraine Severity: [2, nan]
   Nausea: [Yes, No]
   Migraine Start Time: [12:00, nan]
   Light Sensitivity: [Yes, No]
   Migraine End Time: [18:00, nan]
What columns to use for migraine?: Migraine

Given data for columns {data columns}
data:
{data}
what columns to use for {concept}?
```

Figure 5.16: LLM prompts for mapping concepts to relevant columns of data. The text in blue would be filled programmatically. 'IGtype' has the information about the type of the information goal the Analyticon object addresses. 'IGsubtype' specified the instance of the information goal the Analyticon supports.

Making Adjustments

Some Analyticons incorporate special operations common to the information goals and their analytics constructs (**DG4**). These include temporal aggregations and alignments. The former is relevant to Analyticons representing variations over time and is parameterized by the unit of aggregation. The default is month with additional options for weekday, day of month, week of month, and season. These options come from participant needs in MigraineTracker study. Temporal alignment is relevant to Analyticons linking symptoms and treatments or contributors and has multiple types. The default is the same-day alignment with no parameters,

```
You are a data analysis assistant. You are asked to assign a value to {concept} to answer
{question}. Output the value only.

Examples
given [{colum1_1}: {value1_1} | {colum2_1}: {value2_1} | ... | {columN_1}], what's the value of
{concept}?: {valueN_1}
given [{colum1_2}: {value1_2} | {colum2_2}: {value2_2} | ... | {columN_2}], what's the value of
{concept}?: {valueN_2}
given [{colum1_3}: {value1_3} | {colum2_3}: {value2_3} | ... | {columN_3}], what's the value of
{concept}?: {valueN_3}
...

given [{colum1_}: {value1_} | {colum2_}: {value2_} | ... | {columN_}], what's the value of
{concept}?
```

Figure 5.17: LLM prompts for generating values given examples. The text in blue would be filled programmatically.

which considers the co-occurrence of the two variables. Other types are lagged (one variable occurring within n days of the other), cumulative (one variable occurring when n or more days of the other happen consecutively), and let-down (one variable occurring after n or more consecutive days of the other is followed by its absence). These are all parameterized by the number of days (i.e., n), allowing individuals the flexibility of considering various options in the data rendered in the visualizations.

5.3 Representative Analyses

Analyticons support a wide range of analysis, despite *high between and within variations* in information goals as well as the structure and semantics of data, which arise due to individual differences or goal evolution. Individuals can define *concepts over one or more rows and columns*, as needed for their goals. They can also transform values with *simple interactions*. This section illustrates these capabilities.

5.3.1 How Long Are My Migraines?

We demonstrated that Analyticons successfully supported analysis for different ways of recording duration in Section 5.2.1. In the example, Analyticons enabled analysis despite within variations

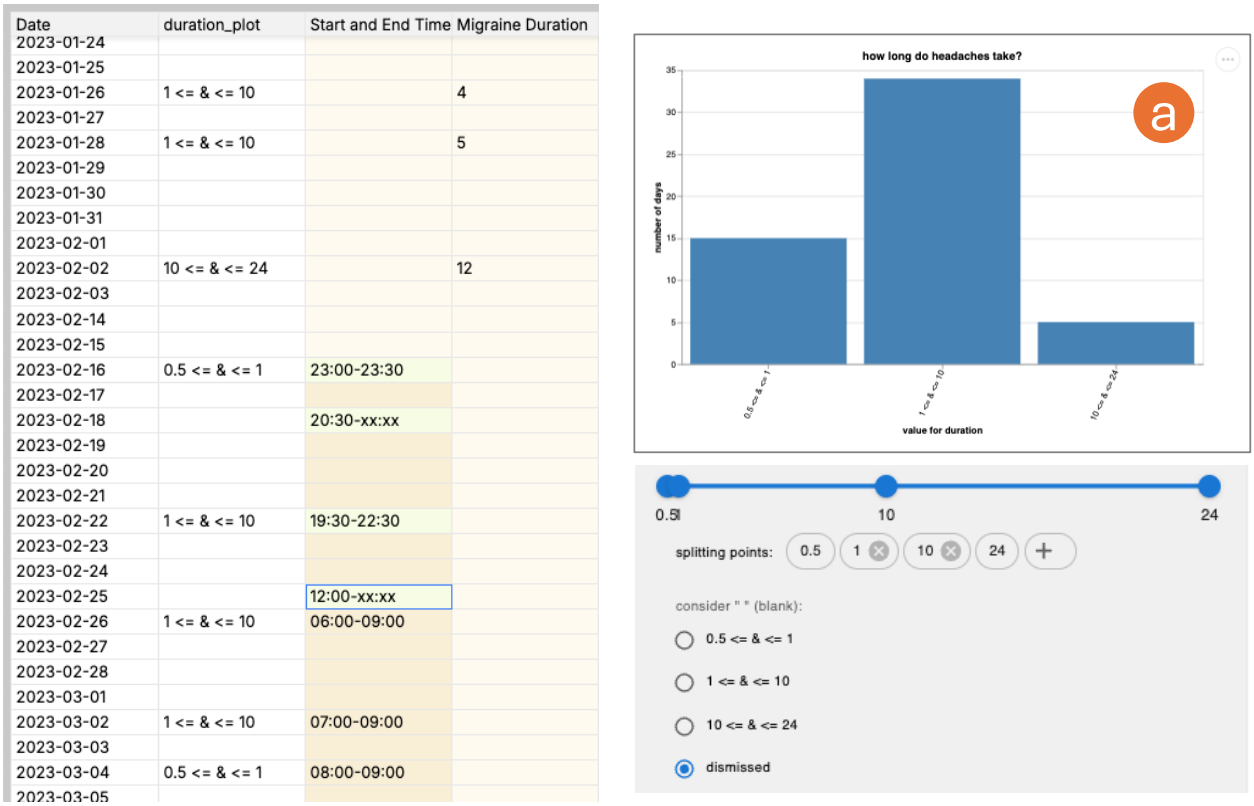


Figure 5.18: Analyticons supported PT10, who recorded duration information as hours as well as start and end times in separate columns. She was able to bin the information according to her needs and different from PT08 in Figure 5.19.

due to goal evolution, where the person changed recording from one column to another and reported duration as a time range rather than hours. Figures 5.18 and 5.19 show handling similar between variations (i.e., variations across individuals). Note that preferences varied for levels of duration to visualize (Figure 5.18 vs. Figure 5.19).

People can approach the same question differently. For example, they can look at duration values over time by choosing a different Analyticon (Figure 5.20). They can also consider a different concept of duration: episode durations. Running with the same example of Section 5.2.1, the person can demonstrate the concept of episode length over consecutive rows of migraine presence then plot this new concept (Figure 5.21). Note that both the columns with relevant information and the concept valuation are very different from the original analysis.

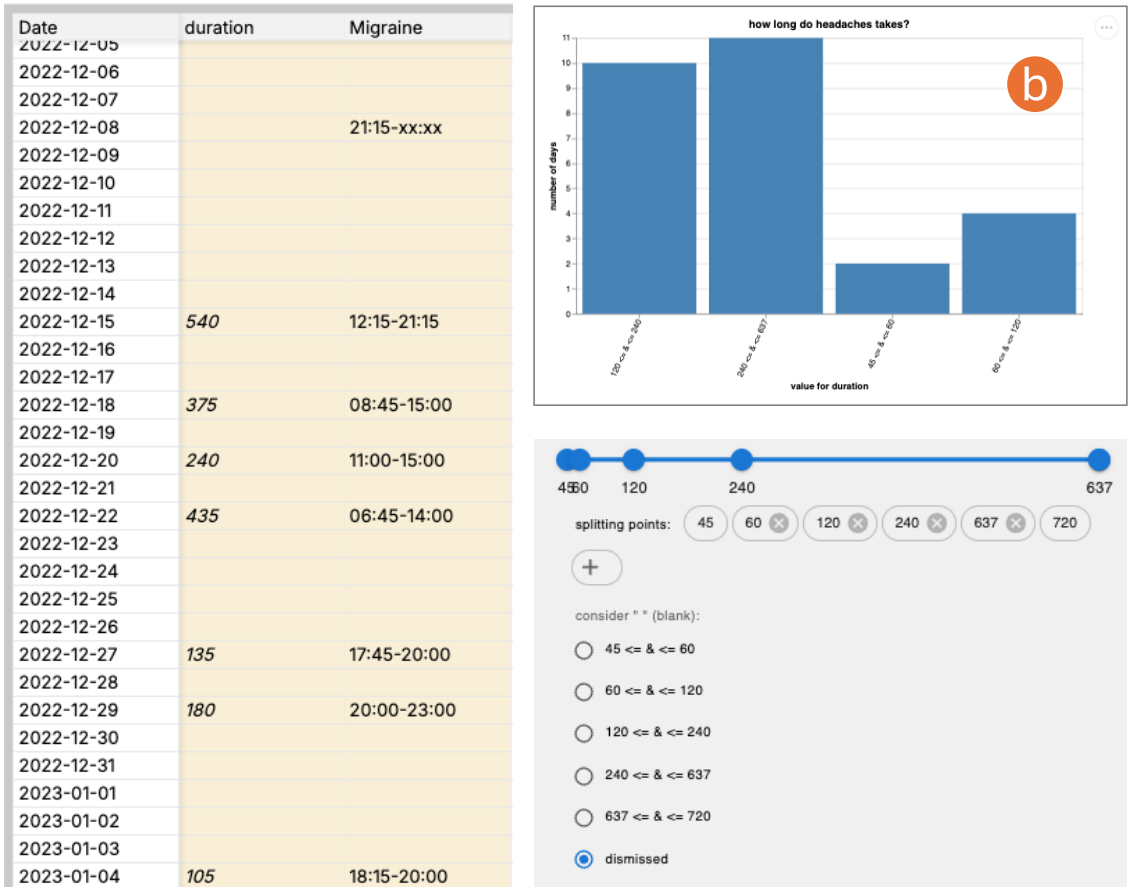
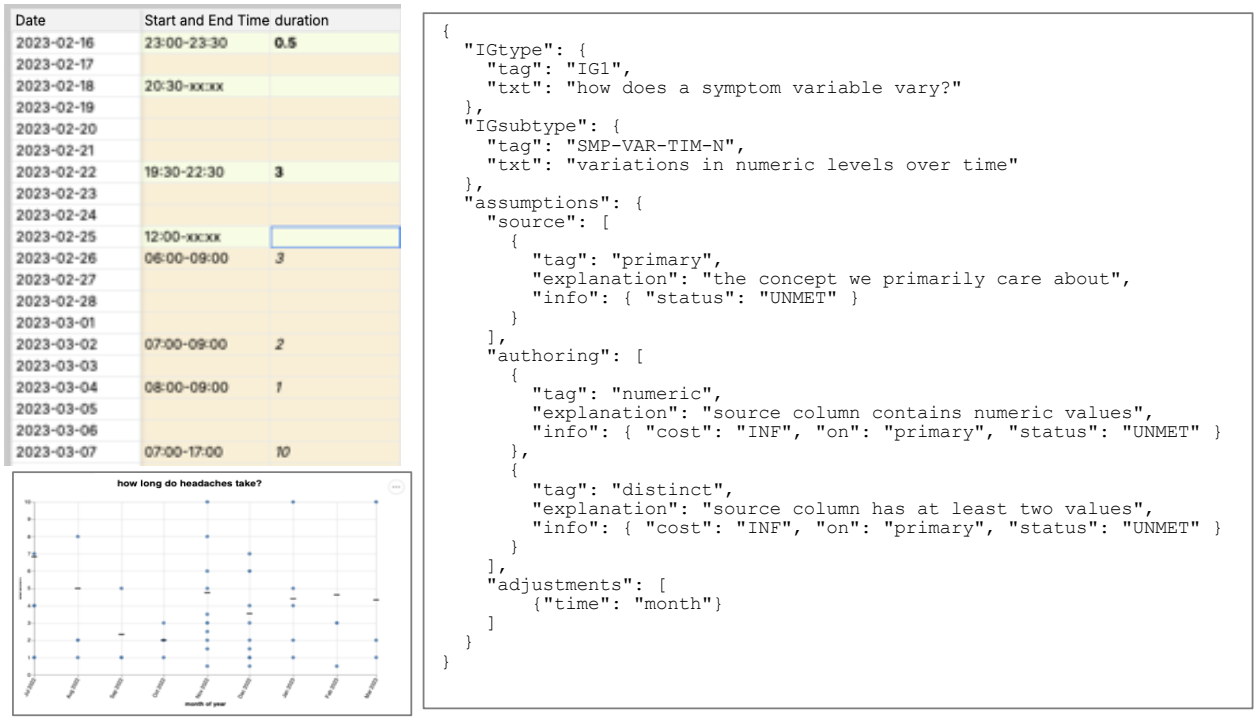


Figure 5.19: Analyticons supported PT08, who recorded duration information as time range under 'Migriane' column. She was interested in binning duration differently compared to PT10 in Figure 5.18.

5.3.2 How Bad Are My Migraines?

Analysis of severity is possible through a very similar set of Analyticons. Although concepts are different, the same analysis applies. Because Analyticons are defined over concepts, they can handle the analysis. We thus use this section to highlight a different aspect of Analyticons: they offer simple interactions so individuals can prepare their data consistent with their knowledge and needs. We briefly showcased one such interaction above in transforming numeric values to ordinal categories using adjustable sliders. Individuals can also dismiss parts of their data as they see fit (Figure 5.22). This was desired in several ways when working with severity



Date	Start and End Time	duration
2023-02-16	23:00-23:30	0.5
2023-02-17		
2023-02-18	20:30-xxxx	
2023-02-19		
2023-02-20		
2023-02-21		
2023-02-22	19:30-22:30	3
2023-02-23		
2023-02-24		
2023-02-25	12:00-xxxx	
2023-02-26	06:00-09:00	3
2023-02-27		
2023-02-28		
2023-03-01		
2023-03-02	07:00-09:00	2
2023-03-03		
2023-03-04	08:00-09:00	1
2023-03-05		
2023-03-06		
2023-03-07	07:00-17:00	10


```
{
  "IGtype": {
    "tag": "IG1",
    "txt": "how does a symptom variable vary?"
  },
  "IGsubtype": {
    "tag": "SMP-VAR-TIM-N",
    "txt": "variations in numeric levels over time"
  },
  "assumptions": {
    "source": [
      {
        "tag": "primary",
        "explanation": "the concept we primarily care about",
        "info": { "status": "UNMET" }
      }
    ],
    "authoring": [
      {
        "tag": "numeric",
        "explanation": "source column contains numeric values",
        "info": { "cost": "INF", "on": "primary", "status": "UNMET" }
      },
      {
        "tag": "distinct",
        "explanation": "source column has at least two values",
        "info": { "cost": "INF", "on": "primary", "status": "UNMET" }
      }
    ],
    "adjustments": [
        {"time": "month"}
    ]
  }
}
```

Figure 5.20: Representation of PT10's data as reports of duration over time. The Analyticon object that enabled this representations appears on the right.

information. Patients sometimes recorded the severity of any headache but considered reports above a threshold to be migraines. They can change what records are considered as migraines by filtering out those below a certain value (Figure 5.22, a). People can also have their individualized outlier removal as was the case for PT06, who removed reports of 0 as outliers (Figure 5.22, b).

5.3.3 Where Do My Migraines Hurt?

In addition to introducing a new Analyticon object, this example shows Analyticons support goal evolution when recording changes from one column to multiple other columns. Analyticons enable this analysis by allowing concept valuation in each period independent of others and by enabling flexible value specification, including those over multiple columns (Figure 5.23).

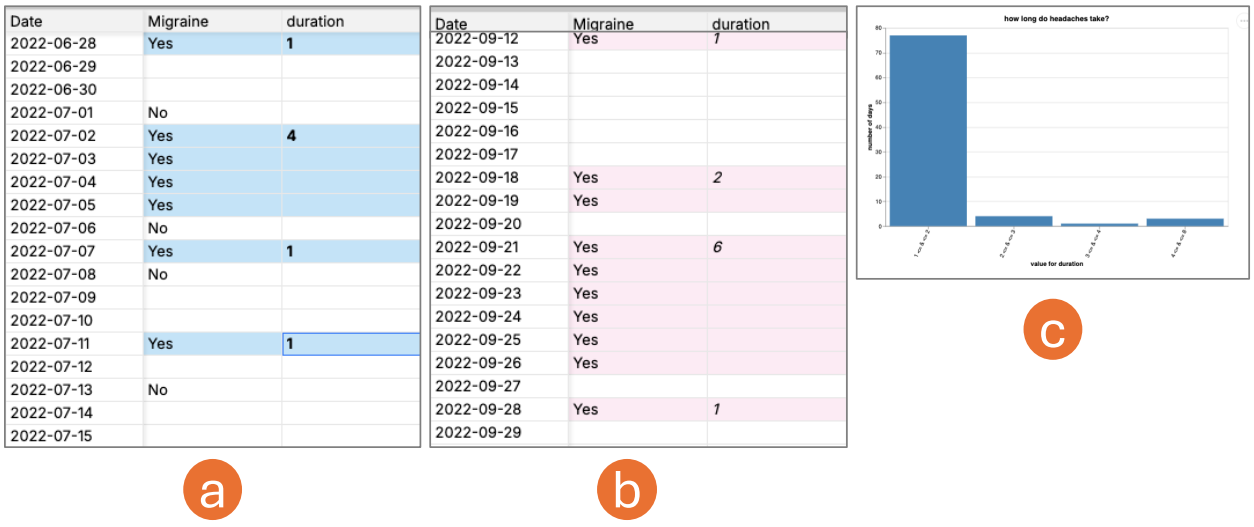


Figure 5.21: It is possible to consider duration of migraine episodes instead of daily migraine durations. PT10 can use the same analysis process for this new information goal. (a) examples demonstrating episode length from reports of migraine presence. (b) automatically generated values for episode duration. Note the ability to define multi-row concept values. (c) new visualization of migraine episode duration.

5.3.4 How Frequent Are My Migraines?

Figure 5.24 shows a new Analyticon and demonstrates the case of handling goal evolution when recording changes within one column: migraine presence reports change from 'Yes'/'No' with occasional missing data to 'Yes'/blank where blank means no migraines. Leveraging her knowledge of when she decided to report differently, the person can use simple interactions to disambiguate blank entries (Figure 5.24, a). She can also adjust the temporal aggregation parameter to her liking (Figure 5.24, b).

5.4 Discussion

We examined data analysis needs and process in working with heterogeneous personal data to address a variety of information goals in the context of chronic conditions. We identified key analytics constructs and the relevant design goals that informed Analyticons, an end-user data

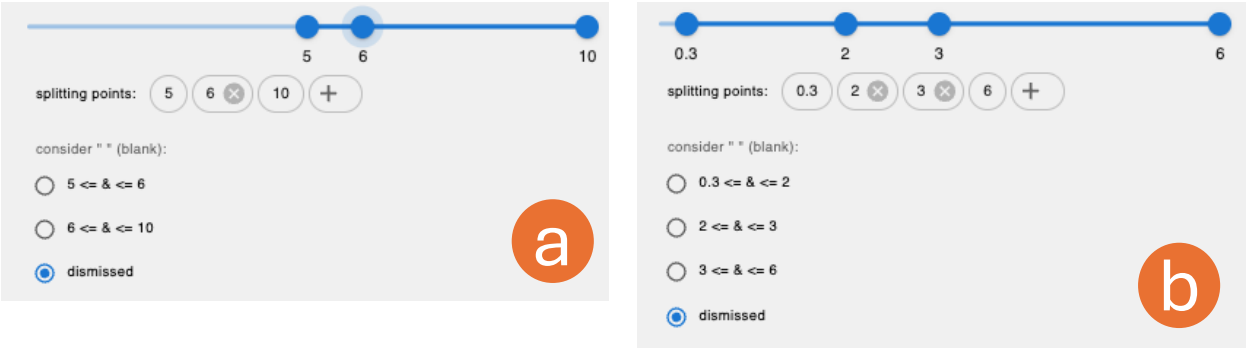


Figure 5.22: Interactive widget enable individualized support for such task as data filtering and outlier removal. (a) PT05 wants to only considers reports of pain above 5 as migraines. (b) PT06 considers reports of zero for migraine severity as outliers.

analysis architecture that represents analysis as constraint satisfaction and adjustments over concepts. Instantiating this architecture with a system powered by LLMs, we demonstrated Analyticons utility in transforming data, step-by-step, from where it is to where it should be to generate goal-aligned visualizations despite high between and within variations in data and information goals. This was achieved by connecting information needs to concepts and visualizations, connecting concepts and visualizations, and leveraging the specific requirements of those visualizations to constrain the data as it relates to the concepts. Exposing the constraints to individuals would allow them to bring in their understanding of the the data and adapt the analysis to their needs. Analyticons system thus combines individual expertise in their own data with analytics expertise as embedded in Analyticon objects and system workflow, and as powered by LLMs. As a result, individuals would have control over the analysis, a capability other research has emphasized (e.g., [82]). Moreover, scaffolding of people's efforts comes as guidance over the steps to take as well as the support in taking them, going beyond what has often been the focus of data analysis tools (e.g., [93, 192, 115]). Below, we further consider Analyticons system capabilities and implications of expressing analysis in terms of concepts, constraints, and adjustments (e.g., the new capabilities it makes possible, Section 5.4.1), point to the value of a domain-specific understanding of data analysis in the design and evaluation

Figure 5.23: PT04's recording of migraine pain location changed from free-form text to reporting 'Yes' under designated location columns. He can obtain location values despite this one-to-many evolution in the recording. Analyticons can handle concepts defined over multiple columns.

of personal informatics tools (Section 5.4.2), and highlight ways Analyticons approach can be extended (Section 5.4.3). We also reflect on the limitations of Analyticons (Section 5.4.4).

5.4.1 Concepts and Constraints Represent End-User Data Analysis

In Analyticons, we explored a novel representation of the analysis process in terms of concepts, constraints, and adjustments. This representation unlocked important capabilities (e.g., data designation, disambiguation, and transformation) in handling high variability in data structures and semantics as well as information goals. Individuals had the ability to adapt the same underlying analysis to a variety of concepts and could do so despite goal evolution. Overall, this representation supported effective goal-directed design. It enabled affordances for the key requirements of evolution and alignment [174]. We further consider the value of

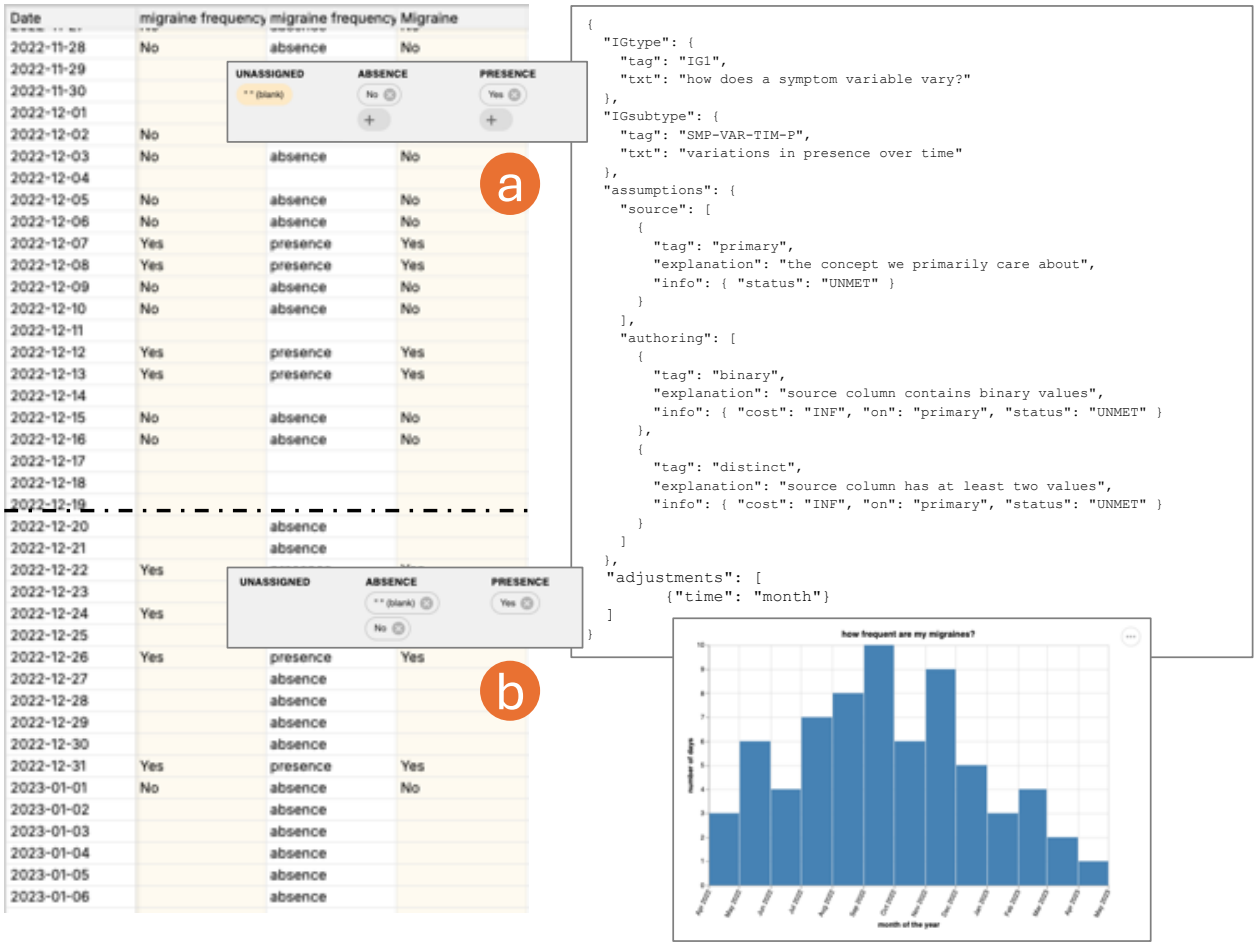


Figure 5.24: PT06 can handle different semantics for blanks. (a) she dismisses blanks as missing data before '2022-12-20'. (b) she considers blank as absence of migraine after '2022-12-22'. She can view the frequency of her migraines using the relevant Analyticon.

this representation: how its structuring of analysis around concepts and constraints made Analyticons possible and what additional opportunities it points to.

Concepts are at a higher level of abstraction than data and disentangle the analysis process from low-level variations of the data. In combination with constraints and adjustment, which allow actions on the concepts, they make it possible to offer more end-user friendly techniques in dealing with complex and tedious tasks (e.g., imputation, outlier handling, slicing, splitting) or to hide them altogether (e.g., aggregation, temporal alignment). This was illustrated in the code-free

interactions of Analyticons for a range of common personal data analysis tasks. In addition to providing useful abstractions, concepts are mediums for higher user involvement and control, so they can align the system with specific needs. These benefits can form the basis for supporting other capabilities. For example, concepts can be leveraged in collaborative analysis settings with clinicians where they bring in their expertise in refining the concepts and thus the analysis. Or, tools can include concept-level error checking based on common ground or medical knowledge to help individuals address data entry errors (e.g., typo entering 10 instead of 1 for medication). In both cases, concepts help leveraging other forms of expertise.

Use of constraints along with a variety of elements has long been appealing in building end-user tools. The most successful examples are the use of geometric constraints over interface elements for layout purposes (e.g., as in Sketchpad [185]) and the use of equation constraints over data elements in spreadsheets [143]. Analyticons explored the applicability of constraints for data analysis purposes over a different type of elements (i.e., goal-aligned concepts). This technique proved useful for the specific and limited domain of our exploration, with similar benefits of enhanced user involvement and higher degrees of control reported elsewhere (e.g., [85]). Our use of constraints was simple and there are other opportunities to leverage them to bring additional analytics expertise. For example, their declarative nature makes them a reasonable avenue for such tasks as guiding individuals through examining internal record inconsistencies and addressing them.

In stepping back and considering how Analyticons were informed by a careful examination of domain-specific analysis tasks for personal data, we can consider the representation of analysis based on concepts and constraints as a way of formalizing domain-specific analysis knowledge for broad application within the domain. Constraints and other constructs within Analyticon objects highlight building blocks for different analysis. As Analyticons are expanded to cover other analysis, these building blocks can become part of a domain specific language for expressing analysis, paralleling efforts in adjacent areas (e.g., visualization authoring) but doing so for analytics needs of a domain.

5.4.2 Goal-Based Characterization of Data Analysis as a Evaluation and Translation Resource

The categories of information goals and the analytics constructs relevant in supporting them provide key requirements to consider in building new personal data tools. This knowledge points to areas for further design and research exploration and offers a principled approach to effectively leveraging emerging technologies (e.g., LLMs). Analyticons illustrated one way of building up on this knowledge but designers can examine other scaffolding techniques for specific goals or analytics constructs.

Past work on design resources for health system design emphasizes the need for making research recommendations accessible to designers in decision-making across contexts and health conditions [106]. In this spirit, the information goals and constructs we identified can become part of a translational resource, particularly in designing for reflection and in considering what and how individuals can learn from their data. Such a resource can encourage designers to take into account the relevant information needs and make research-backed decisions. For example, designers can review the list of constructs and choose the most relevant. A windowing construct is less relevant to a design for self-experimentation, whereas a thresholding construct is more relevant and should receive additional attention in the design.

This resource can also be used to examine how well existing tools support analysis and in what ways they can be improved. More broadly, the information needs we surfaced should be part of evaluating tools for insight discovery. With the increasing interest in automatic or AI-augmented insight discovery, it is particularly important to consider our findings in curating evaluation benchmarks, including those for LLM agents. For example, efforts for building benchmarks for agents specifically designed to support personal health data exploration (e.g., [128]) should be complemented by our findings to ensure these agents can address inquiries people commonly ask in chronic condition management.

5.4.3 Extending and Expanding Analyticons

We explored Analyticons architecture with one of many potential instantiations of it. As an architecture, Analyticons offers both compositionality and extensibility. As to the former, other implementations can be used for exposing and satisfying constraints. These include new widgets and interactions around them, inference engines other than LLMs, or different interaction models. For example, instead of this step-by-step interaction model, one can use a 'repair' model where the system performs the first round of analysis then guides individuals in correcting wherever its assumptions and inferences mismatch that of the individual needs and knowledge. This model would require innovation in scaffolding techniques to guide users on 'where to go back' as well as 'how to make fixes'. As to extensibility of the architecture, additional Analyticon objects can be used for analysis so long as they express the analysis as concepts, constraints, and adjustments. Ideally, we can imagine a store of Analyticons where individuals can browse and obtain new packages of analysis for their data. Such a store can make recommendations given the individual's information goal and then allow them to download the Analyticon object to plug and play the relevant analysis in their analysis system.

The utility of Analyticons for simple isolated analysis flows provides the confidence to further examine whether and how it can enable multi-analysis flows and iterative authoring, dashboard creation, inclusion of additional analysis expertise (e.g., statistical analysis) or visualization support (e.g., by leveraging approaches like DynaVis [186]). These are important open problems to address in designing personal informatics tools that are grounded in people's experiences. As we so far explored with Analyticons, representing analysis process around concepts and constraints can improve support. It can additionally point to areas for further support. For example, Analyticons highlighted points of interaction where powerful computation could augment individuals without compromising their agency. We expect Analyticons to similarly lead to effective ways of leveraging LLMs and other forms of computation to scaffold people's tasks.

5.4.4 Limitations

We supported a wide though incomprehensive set of analyses and used simple implementations under the hood. This was partly because our efforts were focused on *exploring and demonstrating* the core idea rather than *rigorously investigating and optimizing* it. The former informs the latter. This work should thus be considered a sketch highlighting what can be accomplished and not how it can be accomplished [68]. Moreover, we emphasize that our claim is not that any and all analysis can be expressed within this architecture. It is important to establish the expressive power of this technique. Finally, we created each Analyticon based on design goals and through an iterative process of applying them to different cases and improving the designs. As our understanding of analysis tasks in the domain grows, it is also important to investigate more principled ways of developing the objects themselves.

5.5 Contributions to Thesis

In this chapter I introduced Analyticons for data analysis within the goal-directed design framework. Analyticons is an architecture based on constraint-embedded visual objects that enable end-users to analyze their heterogeneous personal data for diverse information needs. Representing analysis through concepts, constraints, and adjustments, it provides means of handling high between and within variability that would allow individuals to align data analysis to their diverse and evolving needs. Analyticons combine system expertise and user expertise: the system offers various forms of scaffolding to guide individuals through the analysis process while individuals can use their unique knowledge and understanding of the data and their needs to control and adjust system behavior.

Chapter 6

DISCUSSION

I demonstrated goal-directed design of personal informatics tools can further realize the potential of personal data for health. MigraineTracker, a novel tool designed around explicit representation of goals, showed that goal expressions and structuring the experience around these expressions can improve planning, collecting, analyzing, interpreting, and acting on personal data. I also identified key requirements for successful goal-directed design: as goals become the first-class design material, we need to support *alignment* of different types of goals with each other and with the system functionality, and in doing so we need to account for *goal evolution* in direct relation with *reflection* on the experiences and the goals. I explored techniques for addressing these requirements in WoNoB and Analyticons. WoNoB leveraged self-regulation techniques to support reflection in planning goal-directed actions under competing demands and changing contexts. Analyticons introduced an architecture for goal-aligned data analysis that accounts for goal evolution. These projects collectively support my thesis statement: *personal health informatics systems with explicit representations of an individual's goals can scaffold planning, collecting, analyzing, interpreting, and acting upon personal data.* Below, I reflect on this work to highlight implications of goal-directed design in addressing a fundamental challenge of HCI: making systems that people with diverse needs can use in different circumstances. I also point to open design problems for future work.

6.1 Avenues for Supporting and Advancing Goal Representations

System functionality of MigraineTracker, WoNoB, and Analyticons was built around explicit representations of goals. However, not all goal types were equally represented across these systems. MigraineTracker allowed individuals to choose such goals as monitoring symptoms

or learning about relations between symptoms and treatments or contributors. These goals corresponded to the ultimate management goal of improving symptoms. Patients then aligned relevant information and tracking goals with these management goals as they configured the app and reviewed their data. Management and tracking goals were explicitly included within the app: management goals as specific selections from a list and tracking goals as app configuration. Although information goals guided preparation of data summaries and visualizations, they were less explicit within the design. WoNoB asked individuals to express work-nonwork balance management goals that were specific, measurable, attainable, relevant, and time-bound (i.e., SMART) then offered implementation intention and mental contrasting guidance around high level information goals regarding an individual's current and desired work-nonwork balance state and opportunities for actions. Tracking goals were minimally represented as the set of activities and whereabouts the individual could report. Analyticons leveraged the information goals in guiding individuals through the analysis process. It did not explicitly include tracking or management goals.

The differences across these systems highlight the importance of examining specialized solutions that are better suited for different goal types. They also point to the need for devising seamless navigation among these solutions as goals are aligned across different types/classes. For instance, we need solutions to better map various devices and sources of data to different tracking goals, tracking models, and data models (e.g., individuals should be able to easily incorporate data from phone sensing or a Fitbit for the desired tracking goal of sleep duration. Similarly, they should be able to change their source of data with ease, or combine data sources). Expanding on prior work (e.g., [105]), we need a new layer of abstraction to enable such interactions. This additional layer provides a shared representation for plugging and playing different data sources without affecting downstream information or management goals and systems built around them (e.g., Analyticons should work with Fitbit sleep data as seamlessly as self-reported sleep data from MigraineTracker). Navigating among solutions and connecting them with each other needs support too. For example, how to take insights surfaced by Analyticons directly back to MigraineTracker and its configuration? What interactions

are relevant? Integration and alignment with external systems is generally important. As people pursue a management goal, how can they adjust the behavior of other applications to align with that goal? For example, how to change the Calendar app around meetings to improve work-nonwork balance? Or, how to adjust Apple's Fitness app to encourage slowing down when physical activity triggers symptoms? As with abstracting away across devices and data sources described above, we need shared and intermediate representations to enable cross-application communication and alignment around goals.

The three systems I introduced were all early stage probes into goal-directed design and thus considered goals within limited scopes. It is, however, important to more deeply consider goals and bring the wealth of decades of research in psychology and HCI in how goals are represented and used ([9, 56, 199]). For example, there are hierarchies within each class of goals, particularly management and information goals (e.g., improving symptoms by first monitoring them and learning their relation with other factors). What would it mean to consider these hierarchies in goal representations? How would those hierarchies influence alignment of different goal types? Goals are also characterized by several dimensions (e.g., importance, difficulty, or temporal range). It is important to incorporate these characteristics in the way goals are represented and used in personal informatics systems.

6.2 Goal-Directed Design to Enable Control and Alignment

As individuals approach a system, including personal data systems, they bring a range of goals. The common practice is for designers to assume certain goals then build the systems to address those goals. In this approach, goals are implicit in the design. With no explicit representation of goals, each system is oblivious to the broader space of goals. It primarily supports the goals it was originally designed for. People who have and maintain the same goal receive support. Others who have different goals or whose goals change cannot adapt and appropriate the system. In the absence of goal representations, systems remain rigid; they do not have the means to offer alternatives (Figure 6.1, a). In contrast, goal-directed design approaches design through goals and their explicit representation. In this approach, goal representations are used

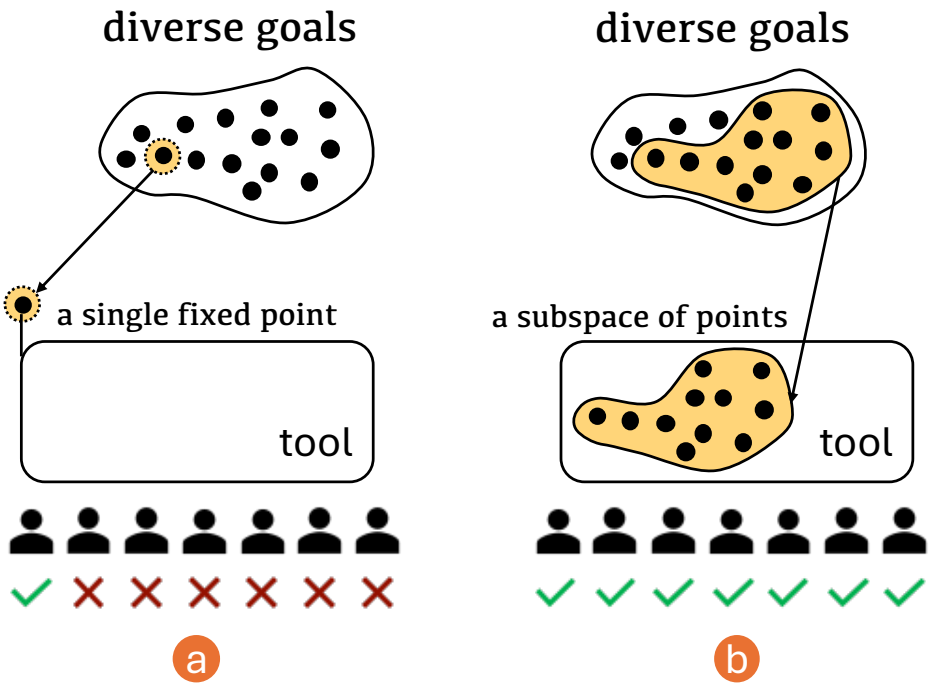


Figure 6.1: Common design practice (a) and goal-directed design (b) in their representation of goals and the role designers play. Common designs do not often include explicit representation of goals. Designers make assumptions about goals and embed in the system. Users cannot pursue other goals. Goal-directed design builds on explicit goal representations to allow users to control the systems and align them to their needs. Designers offer means for users to choose and adjust their goals. Goal-directed design thus provides support to more users.

to structure the experience and system functionality. Rather than being oblivious to the goal space, systems capture a subspace of goals (Figure 6.1, b). Leveraging goal representations, designers can empower individuals to control the system functionality and align it to their diverse needs. Systems no longer correspond to limited and fixed points of the goal space. Each individual decides what points in space a system presents an any given time for their needs. They can also navigate between them, all of which draws on explicit goal representations. A designer's role in goal-directed design is to create representations of the goal subspace and devise control and alignment mechanisms that endow individuals with the ability to pick

relevant goals and navigate between them. I recognize that any system captures part of the goal space. Goal-directed design makes that subspace explicit and thus highlights where additional appropriations and expansions are needed (e.g., by extending goal representations and how they drive system functionality).

6.3 Goal-Directed Design to Support Health Informatics

Goal-directed designs give individuals the ability to appropriate and adjust tools to their health needs. It can also enable a number of other much-needed capabilities in the health context. As a framework, it can guide various aspects of patient support (e.g., in how such support is structured). One example is patient coaching. We can use goals to decide where coaching is needed and what it needs to achieve, then explore techniques (e.g., using AI) for delivering it. Moreover, we can explore how goals can become a *shared representation* to facilitate navigation and information sharing across apps so that personal data activities will no longer be siloed in separate and disjoint apps with poor integration (similar observations discussed in Section 6.1). This can be very helpful in the context of chronic conditions, where comorbidities are common. Currently, individuals are often forced to use different apps that are disconnected, even though their conditions are connected. If designed around goals, information for one condition will be available in considering others. No duplication of efforts and no artificial separation of concerns. This would involve exploring ways goals can enable better data management, sharing, and access while maintaining security and privacy. This spans the use of different platforms (e.g., phone vs. desktop) across care settings (home, work, clinic). With goal-directed design, we can also account for the multi-stakeholder context of care and facilitate communication and collaboration among stakeholder (e.g., clinicians, caregivers, family). For example, we can examine how goals can enable collaborative data analysis in patient-clinician interaction or if they can facilitate experience sharing among patients.

6.4 Goal-Directed Design as a Malleable System Design Approach

It has long been recognized that designers cannot create a design that comes with pre-built functionality to meet the many diverse needs individuals bring to a system [94] and that systems should instead allow individuals to adapt and appropriate them to their needs [58]. Past work on malleable interfaces introduces different methods and techniques for achieving this objective (e.g., [102, 94]). This work often relies on creating novel representations that form new *conceptual models* for building adaptable functionality. For example, one can represent overview-detail design pattern along key dimensions of layout, composition, and content, then enable individuals to manipulate any instance of overview-detail along those dimensions to allow malleability [130]. My work on goal-directed design adds to this body of work. Goals introduce a new layer of abstraction, which in turn provides a basis for mechanisms of adaptation and appropriation as I demonstrated in the previous chapters. They provide a framing to guide where to seek new representations and what capabilities those representations should enable. This way of approaching the creation of representations is particularly suited for such settings as personal data where activities and the accompanying interactions are not as fully-understood as interface design patterns. In these settings goal-directed design can help us identify the fundamental building blocks of the systems and explore the space of interactions to express and combine these blocks.

6.5 Conclusion

I argued that designing health informatics tools around explicit representations of individual goals can help address challenges in planning, collecting, analyzing, interpreting, and acting upon personal data. I supported this argument by introducing three systems. In MigraineTracker, I showed tools which elicit an individual's goals can support highly personalized tracking (e.g., what, when, and how to track) and sense-making (e.g., what data to present and how). I also identified key requirements of goal-centered design: enabling alignment of goals and system functionality, accounting for goal evolution, and facilitating

reflection. In WoNoB, I explored the design for reflection. I leveraged goal pursuit techniques to enable reflection on personal data toward effective actions, despite changing contexts and competing needs. In Analyticons, I explored the design for alignment and evolution with a focus on data analysis. I introduced constraint-embedded goal-aligned visual objects that scaffold end-user analysis of heterogeneous personal data. My work demonstrates the importance of accounting for goals in designing tools that meet the diverse needs their users bring to them. It also points to directions for exploring additional capabilities that support decisions and workflows in selecting, obtaining, interpreting, and acting on data.

BIBLIOGRAPHY

[1] Charles Abraham and Susan Michie. A taxonomy of behavior change techniques used in interventions. *Health psychology*, 27(3):379, 2008.

[2] Daniel A Adler, Emily Tseng, Khatiya C Moon, John Q Young, John M Kane, Emanuel Moss, David C Mohr, and Tanzeem Choudhury. Burnout and the quantified workplace: Tensions around personal sensing interventions for stress in resident physicians. 2022.

[3] Elena Agapie, Patricia A. Areán, Gary Hsieh, and Sean A. Munson. A longitudinal goal setting model for addressing complex personal problems in mental health. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), nov 2022.

[4] Mendinatou Agbetou and Thierry Adoukonou. Lifestyle modifications for migraine management. *Frontiers in Neurology*, 13:719467, 2022.

[5] Tammy D Allen, Eunae Cho, and Laurenz L Meier. Work–family boundary dynamics. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1):99–121, 2014.

[6] Tammy D Allen and LT Eby. The study of interpersonal relationships: An introduction. *Personal Relationships. The Effect on Employee Attitudes, Behavior, and Well-being*, pages 3–14, 2012.

[7] Bon Adriel Aseniero, Charles Perin, Wesley Willett, Anthony Tang, and Sheelagh Carpendale. Activity river: Visualizing planned and logged personal activities for reflection. In *Proceedings of the 2020 International Conference on Advanced Visual Interfaces*, AVI '20, New York, NY, USA, 2020. Association for Computing Machinery.

[8] Blake E Ashforth, Glen E Kreiner, and Mel Fugate. All in a day's work: Boundaries and micro role transitions. *Academy of Management review*, 25(3):472–491, 2000.

[9] James T Austin and Jeffrey B Vancouver. Goal constructs in psychology: Structure, process, and content. *Psychological bulletin*, 120(3):338, 1996.

[10] Amid Ayobi, Paul Marshall, Anna L. Cox, and Yunan Chen. Quantifying the body and caring for the mind: Self-tracking in multiple sclerosis. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 6889–6901, New York, NY, USA, 2017. Association for Computing Machinery.

[11] Eric P.S. Baumer. Reflective informatics: Conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 585–594, New York, NY, USA, 2015. Association for Computing Machinery.

[12] Ute C. Bayer and Peter M. Gollwitzer. Boosting scholastic test scores by willpower: The role of implementation intentions. *Self and Identity*, 6(1):1–19, 2007.

[13] Marit Bentvelzen, Jasmin Niess, Mikołaj P. Woźniak, and Paweł W. Woźniak. The development and validation of the technology-supported reflection inventory. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[14] Marit Bentvelzen, Paweł W. Woźniak, Pia S.F. Herbes, Evropi Stefanidi, and Jasmin Niess. Revisiting reflection in hci: Four design resources for technologies that support reflection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), mar 2022.

[15] Andrew B.L. Berry, Catherine Y. Lim, Calvin A. Liang, Andrea L. Hartzler, Tad Hirsch, Dawn M. Ferguson, Zoë A. Bermet, and James D. Ralston. Supporting collaborative reflection on personal values and health. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.

[16] Toufique Bharmal, Marc Hassenzahl, and Matthias Laschke. From intentions to successful action: Supporting the creation and realization of implementation intentions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–8, New York, NY, USA, 2020. Association for Computing Machinery.

[17] Nicholas Bloom, James Liang, John Roberts, and Zhichun Jenny Ying. Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218, 2015.

[18] Peter Boersma, Lindsey I Black, and Brian W Ward. Prevalence of multiple chronic conditions among us adults, 2018. *Preventing chronic disease*, 17:E106, 2020.

[19] Yanne Bogaerts, Rein De Cooman, and Sara De Gieter. Getting the work-nonwork interface you are looking for: The relevance of work-nonwork boundary management fit. *Frontiers in Psychology*, 9:1158, 2018.

[20] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[21] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, 2019.

[22] Haidy Brown, Ji Sung Kim, and Sue R Faerman. The influence of societal and organizational culture on the use of work-life balance programs: A comparative analysis of the united states and the republic of korea. *The Social Science Journal*, 58(1):62–76, 2021.

[23] Carrie A Bulger, Russell A Matthews, and Mark E Hoffman. Work and personal life boundary management: Boundary strength, work/personal life balance, and the segmentation-integration continuum. *Journal of occupational health psychology*, 12(4):365, 2007.

[24] Scott A Cambo, Daniel Avrahami, and Matthew L Lee. Breaksense: Combining physiological and location sensing to promote mobility during work-breaks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3595–3607, 2017.

[25] James Earl Carnahan. *The effects of self-monitoring by patients on the control of hypertension*. The University of Arizona, USA, 1973.

[26] Chen Chen, Jane Hoffswell, Shunan Guo, Ryan Rossi, Yeuk-Yin Chan, Fan Du, Eunyee Koh, and Zhicheng Liu. Whatsnext: Guidance-enriched exploratory data analysis with interactive, low-code notebooks. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 209–214, 2023.

[27] Gilad Chen, Stanley M Gully, and Dov Eden. Validation of a new general self-efficacy scale. *Organizational research methods*, 4(1):62–83, 2001.

[28] Eunae Cho. Examining boundaries to understand the impact of covid-19 on vocational behaviors, 2020.

[29] Janghee Cho, Tian Xu, Abigail Zimmermann-Niefield, and Stephen Voida. Reflection in theory and reflection in practice: An exploration of the gaps in reflection support among personal informatics apps. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[30] Eun Kyoung Choe, Bongshin Lee, et al. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications*, 35(4):28–37, 2015.

[31] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. Understanding self-reflection: How people reflect on personal data through visual

data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, page 173–182, New York, NY, USA, 2017. Association for Computing Machinery.

[32] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1143–1152, New York, NY, USA, 2014. Association for Computing Machinery.

[33] Shaan Chopra, Rachael Zehrung, Tamil Arasu Shanmugam, and Eun Kyoung Choe. Living with uncertainty and stigma: Self-experimentation and support-seeking around polycystic ovary syndrome. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[34] Tamlin Conner Christensen, Lisa Feldman Barrett, Eliza Bliss-Moreau, Kirsten Lebo, and Cynthia Kaschub. A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, 4(1):53–78, 2003.

[35] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, and Sean A. Munson. More than telemonitoring: Health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. *Journal of Medical Internet Research*, 17:e203, 2015.

[36] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine Zia, James Fogarty, Julie A. Kientz, and Sean A. Munson. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, page 770–786, New York, NY, USA, 2016. Association for Computing Machinery.

[37] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A. Munson. Identifying and planning for individualized change: Patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(1), mar 2019.

[38] Christine Clar, Katharine D Barnard, Ewen Cummins, Pamela Royle, and Norman Waugh. Self-monitoring of blood glucose in type 2 diabetes: systematic review. *Health Technology Assessment*, 14(12):1–140, 2010.

[39] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. No longer wearing: Investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 647–658, New York, NY, USA, 2015. Association for Computing Machinery.

[40] Dave Cook. The freedom trap: digital nomads and the use of disciplining practices to manage work/leisure boundaries. *Information Technology & Tourism*, 22(3):355–390, 2020.

[41] Cary Cooper and Philip Dewe. Well-being—absenteeism, presenteeism, costs and challenges. *Occupational medicine*, 58(8):522–524, 2008.

[42] Felicia Cordeiro, Daniel A. Epstein, Edison Thomaz, Elizabeth Bales, Arvind K. Jagannathan, Gregory D. Abowd, and James Fogarty. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 1159–1162, New York, NY, USA, 2015. Association for Computing Machinery.

[43] Shanley Corvite, Kat Roemmich, Tillie Rosenberg, and Nazanin Andalibi. Data subjects' perspectives on emotion artificial intelligence use in the workplace: A relational ethics lens. *Proceedings of the ACM on Human-Computer Interaction*, 2022.

[44] Aykut Coşkun and Armağan Karahanoğlu. Data sensemaking in self-tracking: Towards a new generation of self-tracking tools. *International Journal of Human–Computer Interaction*, 39(12):2339–2360, 2023.

[45] Ainslea Cross and David Sheffield. Mental contrasting for health behaviour change: A systematic review and meta-analysis of effects and moderator variables. *Health Psychology Review*, 13(2):209–225, 2019.

[46] Vedant Das Swain, Manikanta D Reddy, Kari Anne Nies, Louis Tay, Munmun De Choudhury, and Gregory D Abowd. Birds of a feather clock together: A study of person-organization fit through latent activity routines. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30, 2019.

[47] I Ruiz de Velasco, N González, Y Etxeberria, and JC Garcia-Monco. Quality of life in migraine patients: A qualitative study. *Cephalalgia*, 23(9):892–900, 2003. PMID: 14616931.

[48] Pithalice Delecta. Work life balance. *International journal of current research*, 3(4):186–189, 2011.

[49] Evangelia Demerouti, Daantje Derks, Lieke L Ten Brummelhuis, and Arnold B Bakker. New ways of working: Impact on working conditions, work–family balance, and well-being. In *The impact of ICT on quality of working life*, pages 123–141. Springer, 2014.

[50] Evangelia Demerouti, Akihito Shimazu, Arnold B Bakker, Kyoko Shimada, and Norito Kawakami. Work-self balance: A longitudinal study on the effects of job demands and resources on personal functioning in japanese working parents. *Work & Stress*, 27(3):223–243, 2013.

[51] Aysegül Dogangün, Michael Schwarz, Katharina Kloppenborg, and Robert Le. An approach to improve physical activity by generating individual implementation intentions. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, page 370–375, New York, NY, USA, 2017. Association for Computing Machinery.

[52] Jane E Dutton and Belle Rose Ragins. *Exploring positive relationships at work: Building a theoretical and research foundation.* Psychology Press, 2017.

[53] Domenico D'Amico and Stewart J Tepper. Prophylaxis of migraine: general principles and patient acceptance. *Neuropsychiatric disease and treatment*, 4(6):1155–1167, 2008.

[54] Jeffrey R Edwards and Nancy P Rothbard. Work and family stress and well-being: An examination of person-environment fit in the work and family domains. *Organizational behavior and human decision processes*, 77(2):85–129, 1999.

[55] Elizabeth Victoria Eikey, Clara Marques Caldeira, Mayara Costa Figueiredo, Yunan Chen, Jessica L Borelli, Melissa Mazmanian, and Kai Zheng. Beyond self-reflection: introducing the concept of rumination in personal informatics. *Personal and Ubiquitous Computing*, 25(3):601–616, 2021.

[56] Tina Ekhtiar, Armağan Karahanoğlu, Rúben Gouveia, and Geke Ludden. Goals for goal setting: A scoping review on personal informatics. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 2625–2641, New York, NY, USA, 2023. Association for Computing Machinery.

[57] Carol Emslie and Kate Hunt. 'live to work'or 'work to live'? a qualitative study of gender and work–life balance among men and women in mid-life. *Gender, Work & Organization*, 16(1):151–172, 2009.

[58] Douglas C Engelbart. Augmenting human intellect: a conceptual framework (1962). 2021.

[59] Daniel A Epstein, Daniel Avrahami, and Jacob T Biehl. Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 673–684, 2016.

[60] Daniel A. Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M. Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, Payam Dowlatyari, Craig Hilby, Sazeda Sultana, Elizabeth V. Eikey, and Yunan Chen. Mapping and taking stock of the personal informatics literature. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(4), December 2020.

[61] Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. Beyond abandonment to next steps: Understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 1109–1113, New York, NY, USA, 2016. Association for Computing Machinery.

[62] Daniel A. Epstein, Nicole B. Lee, Jennifer H. Kang, Elena Agapie, Jessica Schroeder, Laura R. Pina, James Fogarty, Julie A. Kientz, and Sean Munson. Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 6876–6888, New York, NY, USA, 2017. Association for Computing Machinery.

[63] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 731–742, New York, NY, USA, 2015. Association for Computing Machinery.

[64] A. Farmer, O. J. Gibson, L. Tarassenko, and A. Neil. A systematic review of telemedicine interventions to support blood glucose self-monitoring in diabetes. *Diabetic Medicine*, 22(10):1372–1378, 2005.

[65] Michel D Ferrari. The economic burden of migraine to society. *Pharmacoeconomics*, 13:667–676, 1998.

[66] Martin Fishbein and Joseph N Cappella. The role of theory in developing effective health communications. *Journal of communication*, 56(suppl_1):S1–S17, 2006.

[67] Rowanne Fleck and Geraldine Fitzpatrick. Reflecting on reflection: Framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, page 216–223, New York, NY, USA, 2010. Association for Computing Machinery.

[68] James Fogarty. Code and contribution in interactive systems research. In *Workshop HCITools: Strategies and Best Practices for Designing, Evaluating and Sharing Technical HCI Toolkits at CHI*, pages 1–4, 2017.

[69] Annie Foucreault, Ariane Ollier-Malaterre, and Julie Ménard. Organizational culture and work–life integration: A barrier to employees' respite? *The International Journal of Human Resource Management*, 29(16):2378–2398, 2018.

[70] Robert H Friedman, Lewis E Kazis, Alan Jette, Mary Beth Smith, John Stollerman, Jeanne Torgerson, and Kathleen Carey. A telecommunications system for monitoring

and counseling patients with hypertension: impact on medication adherence and blood pressure control. *American journal of hypertension*, 9(4):285–292, 1996.

[71] Paul Glasziou, Les Irwig, and David Mant. Monitoring in chronic disease: a rational approach. *Bmj*, 330(7492):644–648, 2005.

[72] Lewis R Goldberg et al. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7(1):7–28, 1999.

[73] Anton Gollwitzer, Gabriele Oettingen, Teri A Kirby, Angela L Duckworth, and Doris Mayer. Mental contrasting facilitates academic performance in school children. *Motivation and Emotion*, 35(4):403–412, 2011.

[74] Peter M. Gollwitzer and Paschal Sheeran. Implementation intentions and goal achievement: A meta-analysis of effects and processes. 38:69–119, 2006.

[75] Peter M Gollwitzer, Frank Wieber, Andrea L Myers, Sean M McCrea, C Agnew, D Carlston, et al. How to maximize implementation intention effects. *Then a miracle occurs: Focusing on behavior in social psychological theory and research*, pages 137–161, 2010.

[76] Rúben Gouveia and Daniel A. Epstein. This watchface fits with my tattoos: Investigating customisation needs and preferences in personal tracking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[77] Grand Canyon University. Work-Life Balance Stats Among Business Professionals, 2020.

[78] Jeffrey H Greenhaus and Tammy D Allen. Work–family balance: A review and extension of the literature. 2011.

[79] Erik Grönvall and Nervo Verdezoto. Understanding challenges and opportunities of preventive blood pressure self-monitoring at home. In *Proceedings of the 31st European Conference on Cognitive Ergonomics*, ECCE '13, New York, NY, USA, 2013. Association for Computing Machinery.

[80] Joseph G Grzywacz, David M Almeida, and Daniel A McDonald. Work–family spillover and daily reports of work and family stress in the adult labor force. *Family relations*, 51(1):28–36, 2002.

[81] Ken Gu, Ruoxi Shang, Ruien Jiang, Keying Kuang, Richard-John Lin, Donghe Lyu, Yue Mao, Youran Pan, Teng Wu, Jiaqian Yu, et al. Blade: Benchmarking language model agents for data-driven science. *arXiv preprint arXiv:2408.09667*, 2024.

[82] Jiajing Guo, Vikram Mohanty, Jorge H Piazentin Ono, Hongtao Hao, Liang Gou, and Liu Ren. Investigating interaction modes and user agency in human-llm collaboration for domain-specific data analysis. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

[83] Karen Hacker. The burden of chronic disease. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, 8(1):112–119, 2024.

[84] Cother Hajat and Emma Stein. The global burden of multiple chronic conditions: A narrative review. *Preventive Medicine Reports*, 12:284–293, 2018.

[85] Jiawei Han, L.V.S. Lakshmanan, and R.T. Ng. Constraint-based, multidimensional data mining. *Computer*, 32(8):46–50, 1999.

[86] Paul L Hewitt and Gordon L Flett. Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. *Journal of personality and social psychology*, 60(3):456, 1991.

[87] Esther Howe, Jina Suh, Mehrab Bin Morshed, Daniel McDuff, Kael Rowan, Javier Hernandez, Marah Ihab Abdin, Gonzalo Ramos, Tracy Tran, and Mary P Czerwinski. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[88] Dandan Huang, Melanie Tory, and Lyn Bartram. A field study of on-calendar visualizations. In *Proceedings of the 42nd Graphics Interface Conference*, GI '16, page 13–20, Waterloo, CAN, 2016. Canadian Human-Computer Communications Society.

[89] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, page 17–24, New York, NY, USA, 2003. Association for Computing Machinery.

[90] Susan L. Janson, Kelly Wong McGrath, Jack K. Covington, Su-Chun Cheng, and Homer A. Boushey. Individualized asthma self-management improves medication adherence and markers of asthma control. *Journal of Allergy and Clinical Immunology*, 123(4):840–846, 2009.

[91] Matthew Jörke, Yasaman S. Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. Pearl: A technology probe for machine-assisted reflection on personal data. *28th International Conference on Intelligent User Interfaces*, 2023.

[92] Daniel Kahneman, Alan B Krueger, David A Schkade, Norbert Schwarz, and Arthur A Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 2004.

[93] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Wrangler: interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 3363–3372, New York, NY, USA, 2011. Association for Computing Machinery.

[94] David R. Karger. Haystack: Per-user information environments based on semistructured data. In *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*. The MIT Press, 03 2007.

[95] Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. Tummytrials: A feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 6850–6863, New York, NY, USA, 2017. Association for Computing Machinery.

[96] Ravi Karkar, Jasmine Zia, Roger Vilardaga, Sonali R Mishra, James Fogarty, Sean A Munson, and Julie A Kientz. A framework for self-experimentation in personalized health. *Journal of the American Medical Informatics Association*, 23(3):440–448, 2016.

[97] Alboukadel Kassambara. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*, 2018. R package version 0.7.1.

[98] Dmitri S. Katz, Blaine A. Price, Simon Holland, and Nicholas Sheep Dalton. Data, data everywhere, and still too hard to link: Insights from user interactions with diabetes apps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[99] Harmanpreet Kaur, Daniel McDuff, Alex C Williams, Jaime Teevan, and Shamsi T Iqbal. "i didn't know i looked angry": Characterizing observed emotion and reported affect at work. In *CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2022.

[100] Harmanpreet Kaur, Alex C. Williams, Daniel McDuff, Mary Czerwinski, Jaime Teevan, and Shamsi T. Iqbal. Optimizing for happiness and productivity: Modeling opportune moments for transitions and breaks at work. In *Proceedings of the 2020 CHI Conference on Human*

Factors in Computing Systems, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

[101] Anna Kawakami, Shreya Chowdhary, Shamsi T Iqbal, Q Vera Liao, Alexandra Olteanu, Jina Suh, and Koustuv Saha. Sensing wellbeing in the workplace, why and for whom? envisioning impacts with organizational stakeholders. *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, 2023.

[102] Alan C. Kay. The early history of smalltalk. In *The Second ACM SIGPLAN Conference on History of Programming Languages*, HOPL-II, page 69–95, New York, NY, USA, 1993. Association for Computing Machinery.

[103] Matthew Kay. Challenges in personal health tracking: the data isn't enough. *XRDS*, 21(2):32–37, dec 2014.

[104] L Kelman. The triggers or precipitants of the acute migraine attack. *Cephalalgia*, 27(5):394–402, 2007. PMID: 17403039.

[105] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. Omnitrack: A flexible self-tracking approach leveraging semi-automated tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3), sep 2017.

[106] Susanne Kirchner, Jessica Schroeder, James Fogarty, and Sean A. Munson. "they don't always think about that": Translational needs in the design of personal health informatics applications. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[107] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. Reflection companion: A conversational system for engaging users in reflection on physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2), jul 2018.

[108] Amanda Lazar, Christian Koehler, Theresa Jean Tanenbaum, and David H. Nguyen. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 635–646, New York, NY, USA, 2015. Association for Computing Machinery.

[109] Min Kyung Lee, Junsung Kim, Jodi Forlizzi, and Sara Kiesler. Personalization revisited: A reflective approach helps people better personalize health services and motivates them to increase physical activity. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 743–754, New York, NY, USA, 2015. Association for Computing Machinery.

[110] Aniek Lentferink, Matthijs L. Noordzij, Anouk Burgler, Randy Klaassen, Youri Derks, Hilbrand Oldenhuis, Hugo Velthuijsen, and Lisette van Gemert-Pijnen. On the receptivity of employees to just-in-time self-tracking and ecoaching for stress management: a mixed-methods approach. *Behaviour & Information Technology*, 41(7):1398–1424, 2022.

[111] Ian Li, Anind Dey, and Jodi Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 557–566, New York, NY, USA, 2010. Association for Computing Machinery.

[112] Ian Li, Anind K. Dey, and Jodi Forlizzi. Using context to reveal factors that affect physical activity. *ACM Trans. Comput.-Hum. Interact.*, 19(1), may 2012.

[113] Peng Liao, Walter Dempsey, Hillol Sarker, Syed Monowar Hossain, Mustafa al'Absi, Predrag Klasnja, and Susan Murphy. Just-in-time but not too much: Determining treatment timing in mobile health. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018.

[114] Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(1), mar 2020.

[115] Michael Xieyang Liu, Advait Sarkar, Carina Negreanu, Benjamin Zorn, Jack Williams, Neil Toronto, and Andrew D. Gordon. "what it wants me to say": Bridging the abstraction gap between end-user programmers and code-generating large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[116] Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. Data illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.

[117] Edwin A Locke and Gary P Latham. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, 57(9):705, 2002.

[118] Edwin A Locke and Gary P Latham. Goal setting theory: The current state. page 623–630, 2013.

[119] Daniela Lohaus and Wolfgang Habermann. Presenteeism: A review and research directions. *Human Resource Management Review*, 29(1):43–58, 2019.

[120] Christopher A Magee, Natalie Stefanic, Peter Caputi, and Don C Iverson. The association between job demands/control and health in employed parents: The mediating role of work-to-family interference and enhancement. *Journal of Occupational Health Psychology*, 17(2):196, 2012.

[121] Lena Mamykina, Elizabeth M Heitkemper, Arlene M Smaldone, Rita Kukafka, Heather Cole-Lewis, Patricia G Davidson, Elizabeth D Mynatt, Jonathan N Tobin, Andrea Cassells, Carrie Goodman, et al. Structured scaffolding for reflection and problem solving in diabetes self-management: qualitative study of mobile diabetes detective. *Journal of the American Medical Informatics Association*, 23(1):129–136, 2016.

[122] Lena Mamykina, Elizabeth M. Heitkemper, Arlene M. Smaldone, Rita Kukafka, Heather J. Cole-Lewis, Patricia G. Davidson, Elizabeth D. Mynatt, Andrea Cassells, Jonathan N. Tobin, and George Hripcsak. Personal discovery in diabetes self-management: Discovering cause and effect using self-monitoring data. *Journal of Biomedical Informatics*, 76:1–8, 2017.

[123] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. Mahi: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, page 477–486, New York, NY, USA, 2008. Association for Computing Machinery.

[124] Aubrey N Manack, Dawn C Buse, and Richard B Lipton. Chronic migraine: epidemiology and disease burden. *Current pain and headache reports*, 15:70–78, 2011.

[125] Michael J. Marmura, Stephen D. Silberstein, and Todd J. Schwedt. The acute treatment of migraine in adults: The american headache society evidence assessment of migraine pharmacotherapies. *Headache: The Journal of Head and Face Pain*, 55(1):3–20, 2015.

[126] Melissa Mazmanian, Wanda J Orlikowski, and JoAnne Yates. The autonomy paradox: The implications of mobile email devices for knowledge professionals. *Organization science*, 24(5):1337–1357, 2013.

[127] Helena M. Mentis, Anita Komlodi, Katrina Schrader, Michael Phipps, Ann Gruber-Baldini, Karen Yarbrough, and Lisa Shulman. Crafting a view of self-tracking data in the clinical visit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 5800–5812, New York, NY, USA, 2017. Association for Computing Machinery.

[128] Mike A Merrill, Akshay Paruchuri, Naghmeh Rezaei, Geza Kovacs, Javier Perez, Yun Liu, Erik Schenck, Nova Hammerquist, Jake Sunshine, Shyam Tailor, et al. Transforming wearable data into health insights using large language model agents. *arXiv preprint arXiv:2406.06464*, 2024.

[129] André N. Meyer, Gail C. Murphy, Thomas Zimmermann, and Thomas Fritz. Enabling good work habits in software developers through reflective goal-setting. *IEEE Transactions on Software Engineering*, 47(9):1872–1885, 2021.

[130] Bryan Min, Allen Chen, Yining Cao, and Haijun Xia. Malleable overview-detail interfaces. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.

[131] Sonali R. Mishra, Predrag Klasnja, John MacDuffie Woodburn, Eric B. Hekler, Larsson Omberg, Michael Kellen, and Lara Mangravite. Supporting coping with parkinson's disease through self tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–16, New York, NY, USA, 2019. Association for Computing Machinery.

[132] Varun Mishra, Florian Künzler, Jan-Niklas Kramer, Elgar Fleisch, Tobias Kowatsch, and David Kotz. Detecting receptivity for mhealth interventions in the natural environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(2), jun 2021.

[133] Jimmy Moore, Pascal Goffin, Jason Wiese, and Miriah Meyer. An interview method for engaging personal data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4), dec 2022.

[134] Sean A. Munson. Rethinking assumptions in the design of health and wellness tracking tools. *Interactions*, 25(1):62–65, dec 2017.

[135] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Voida, and Jaime Snyder. Personal informatics in interpersonal contexts: Towards the design of technology that supports the social ecologies of long-term mental health management. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018.

[136] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6):446–462, 2018.

[137] Jasmin Niess and Paweł W. Woźniak. Supporting meaningful personal fitness: The tracker goal evolution model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA, 2018. Association for Computing Machinery.

[138] Gabriele Oettingen and Peter M Gollwitzer. Health behavior change by self-regulation of goal pursuit: Mental contrasting with implementation intentions. In *The Routledge*

international handbook of self-control in health and well-being, pages 418–430. Routledge, 2017.

[139] Gabriele Oettingen, Heather Barry Kappes, Katie B. Guttenberg, and Peter M. Gollwitzer. Self-regulation of time management: Mental contrasting with implementation intentions. *European Journal of Social Psychology*, 45(2):218–229, 2015.

[140] Gabriele Oettingen, Hyeon-ju Pak, and Karoline Schnetter. Self-regulation of goal-setting: Turning free fantasies about the future into binding goals. *Journal of personality and social psychology*, 80(5):736, 2001.

[141] Gabriele Oettingen and Marion Wittchen. Regulating goal pursuit through mental contrasting with implementation intentions. In *New developments in goal setting and task performance*, pages 547–572. Routledge, 2013.

[142] Julie B Olson-Buchanan and Wendy R Boswell. Blurring boundaries: Correlates of integration and segmentation between work and nonwork. *Journal of Vocational behavior*, 68(3):432–445, 2006.

[143] Stephen W. Oney. *Expressing Interactivity with States and Constraints*. PhD thesis, 2015. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-03-04.

[144] Sheina Orbell and Paschal Sheeran. 'inclined abstainers': A problem for predicting health-related behaviour. *British Journal of Social Psychology*, 37(2):151–165, 1998.

[145] Aisling Ann O'kane, Sun Young Park, Helena Mentis, Ann Blandford, and Yunan Chen. Turning to peers: integrating understanding of the self, the condition, and others' experiences in making sense of complex chronic conditions. *Computer Supported Cooperative Work (CSCW)*, 25:477–501, 2016.

[146] Aditeya Pandey, Arjun Srinivasan, and Vidya Setlur. MEDLEY: Intent-based Recommendations to Support Dashboard Composition . *IEEE Transactions on Visualization & Computer Graphics*, 29(01):1135–1145, January 2023.

[147] Sun Young Park and Yunan Chen. Individual and social recognition: Challenges and opportunities in migraine management. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, page 1540–1551, New York, NY, USA, 2015. Association for Computing Machinery.

[148] James W Peltier, Amanda Hay, and William Drago. The reflective learning continuum: Reflecting on reflection. *Journal of marketing education*, 27(3):250–263, 2005.

[149] Ingrid Pettersson, Florian Lachner, Anna-Katharina Frison, Andreas Riener, and Andreas Butz. A bermuda triangle? a review of method application and triangulation in user experience evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–16, New York, NY, USA, 2018. Association for Computing Machinery.

[150] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. When attention is not scarce - detecting boredom from mobile phone usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 825–836, New York, NY, USA, 2015. Association for Computing Machinery.

[151] Laura R. Pina, Sang-Wha Sien, Teresa Ward, Jason C. Yip, Sean A. Munson, James Fogarty, and Julie A. Kientz. From personal informatics to family informatics: Understanding family practices around health monitoring. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 2300–2315, New York, NY, USA, 2017. Association for Computing Machinery.

[152] Charlie Pinder, Jo Vermeulen, Adhi Wicaksono, Russell Beale, and Robert J. Hendley. If this, then habit: Exploring context-aware implementation intentions on smartphones. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, page 690–697, New York, NY, USA, 2016. Association for Computing Machinery.

[153] Helen Pluut and Jaap Wonders. Not able to lead a healthy life when you need it the most: Dual role of lifestyle behaviors in the association of blurred work-life boundaries with well-being. *Frontiers in psychology*, 11:607294, 2020.

[154] Andrew Prestwich, Paschal Sheeran, Thomas L Webb, and Peter M Gollwitzer. Implementation intentions. *Predicting health behavior*, 3:321–357, 2015.

[155] James O. Prochaska and Wayne F. Velicer. The transtheoretical model of health behavior change. *American journal of health promotion*, 12(1):38–48, 1997.

[156] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.

[157] Sumita Raghuram and Batia Wiesenfeld. Work-nonwork conflict and job stress among virtual workers. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, 43(2-3):259–277, 2004.

[158] Protiva Rahman, Courtney Hebert, and Arnab Nandi. Icarus: minimizing human effort in iterative data completion. In *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*, volume 11, page 2263, 2018.

[159] Shriti Raj, Toshi Gupta, Joyce Lee, Matthew Kay, and Mark W Newman. "it can bring you in the right direction": Episode-driven data narratives to help patients navigate multidimensional diabetes data to make care decisions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[160] Shriti Raj, Mark W. Newman, Joyce M. Lee, and Mark S. Ackerman. Understanding individual and collaborative problem-solving with patient-generated data: Challenges and opportunities. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW), dec 2017.

[161] Amon Rapp and Federica Cena. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, 94:1–17, 2016.

[162] Damien Ridge, Laurna Bullock, Hilary Causer, Tamsin Fisher, Samantha Hider, Tom Kingstone, Lauren Gray, Ruth Riley, Nina Smyth, Victoria Silverwood, Johanna Spiers, and Jane Southam. 'imposter participants' in online qualitative research, a new and increasing threat to data integrity? *Health Expectations*, 26(3):941–944, 2023.

[163] Ivan T Robertson, Cary L Cooper, and Sheena Johnson. *Well-being: Productivity and happiness at work*, volume 3. Springer, 2011.

[164] Kat Roemmich and Nazanin Andalibi. Data subjects' conceptualizations of and attitudes toward automatic emotion recognition-enabled wellbeing interventions on social media. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–34, 2021.

[165] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal tracking as lived informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 1163–1172, New York, NY, USA, 2014. Association for Computing Machinery.

[166] Anna Rudnicka, Dave Cook, Marta E Cecchinato, Sandy JJ Gould, Joseph W Newbold, and Anna L Cox. The end of the active work break? remote work, sedentariness and the role of technology in creating active break-taking norms. In *2022 Symposium on Human-Computer Interaction for Work*, pages 1–13, 2022.

[167] Hyeyoung Ryu, Andrew B.L. Berry, Catherine Y Lim, Andrea Hartzler, Tad Hirsch, Juanita I Trejo, Zoë Abigail Bermet, Brandi Crawford-Gallagher, Vi Tran, Dawn Ferguson, David J

Cronkite, Brooks Tiffany, John Weeks, and James Ralston. "you can see the connections": Facilitating visualization of care priorities in people living with multiple chronic health conditions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[168] Marc Schabracq, Jacques AM Winnubst, Cary L Cooper, et al. *The handbook of work and health psychology*. Wiley Online Library, 2003.

[169] Morgan Klaus Scheuerman, Katta Spiel, Oliver L. Haimson, Foad Hamidi, and Stacy M. Branham. Hci guidelines for gender equity and inclusivity, 2020.

[170] Jessica Schroeder, Chia-Fang Chung, Daniel A. Epstein, Ravi Karkar, Adele Parsons, Natalia Murinova, James Fogarty, and Sean A. Munson. Examining self-tracking by people with migraine: Goals, needs, and opportunities in a chronic health condition. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, page 135–148, New York, NY, USA, 2018. Association for Computing Machinery.

[171] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, page 1726–1739, New York, NY, USA, 2017. Association for Computing Machinery.

[172] Jessica Schroeder, Ravi Karkar, James Fogarty, Julie A Kientz, Sean A Munson, and Matthew Kay. A patient-centered proposal for bayesian analysis of self-experiments for health. *Journal of healthcare informatics research*, 3:124–155, 2019.

[173] Jessica Schroeder, Ravi Karkar, Natalia Murinova, James Fogarty, and Sean A. Munson. Examining opportunities for goal-directed self-tracking to support chronic condition management. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4), sep 2020.

[174] Yasaman S. Sefidgar, Carla L. Castillo, Shaan Chopra, Liwei Jiang, Tae Jones, Anant Mittal, Hyeyoung Ryu, Jessica Schroeder, Allison Cole, Natalia Murinova, Sean A. Munson, and James Fogarty. Migrainetracker: Examining patient experiences with goal-directed self-tracking for a chronic health condition. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[175] Yasaman S. Sefidgar, Matthew Jörke, Jina Suh, Koustuv Saha, Shamsi Iqbal, Gonzalo Ramos, and Mary P Czerwinski. Lessons learned for data-driven implementation intentions with mental contrasting. In *Extended Abstracts of the 2023 CHI Conference on Human Factors*

in Computing Systems, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery.

[176] Yasaman S. Sefidgar, Jina Sun, Jeffrey Heer, Sean Munson, and James Fogarty. Analyticons: an architecture for end-user interactive analysis of personal data. 2025.

[177] Paschal Sheeran. Intention—behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1):1–36, 2002.

[178] Tarique Siddiqui, Albert Kim, John Lee, Karrie Karahalios, and Aditya Parameswaran. Effortless data exploration with zenvisage: an expressive and interactive visual analytics system. *Proc. VLDB Endow.*, 10(4):457–468, November 2016.

[179] Lucas M. Silva, Xi Lu, Emily X. Liang, and Daniel A. Epstein. Foody talk: Exploring opportunities for conversational food journaling. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.

[180] Robert Smith. Impact of migraine on the family. *Headache: The Journal of Head and Face Pain*, 38(6):423–426, 1998.

[181] Glen D Solomon and KL Price. Burden of migraine: a review of its socioeconomic impact. *Pharmacoeconomics*, 11:1–10, 1997.

[182] Alfonso Sousa-Poza and Andres A Sousa-Poza. Well-being at work: a cross-national analysis of the levels and determinants of job satisfaction. *The journal of socio-economics*, 29(6):517–538, 2000.

[183] Arjun Srinivasan and Vidya Setlur. Bolt: A natural language interface for dashboard authoring. In *EuroVis (Short Papers)*, pages 7–11, 2023.

[184] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. Bridging the gulf of envisioning: Cognitive challenges in prompt based interactions with llms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[185] I. E. Sutherland. Sketchpad a man-machine graphical communication system. In *Papers on Twenty-Five Years of Electronic Design Automation*, 25 years of DAC, page 507–524, New York, NY, USA, 1988. Association for Computing Machinery.

[186] Priyan Vaithilingam, Elena L. Glassman, Jeevana Priya Inala, and Chenglong Wang. Dynavis: Dynamically synthesized ui widgets for visualization editing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery.

[187] Stefanie Valentic. More Than Half of Americans Have Unhealthy Work-Life Balance, 2019.

[188] Niels van Berkel and Vassilis Kostakos. Recommendations for conducting longitudinal experience sampling studies. In *Advances in Longitudinal HCI Research*, pages 59–78. Springer, 2021.

[189] Elisabeth T. Kersten van Dijk, Joyce H.D.M. Westerink, Femke Beute, and Wijnand A. IJsselsteijn. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction*, 32(5-6):268–296, 2017.

[190] Lisa M Vizer, Jordan Eschler, Bon Mi Koo, James Ralston, Wanda Pratt, and Sean Munson. "it's not just technology, it's people": Constructing a conceptual model of shared health informatics for tracking in chronic illness management. *J Med Internet Res*, 21(4):e10830, Apr 2019.

[191] Chenglong Wang, Bongshin Lee, Steven M. Drucker, Dan Marshall, and Jianfeng Gao. Data formulator 2: Iterative creation of data visualizations, with ai transforming data along the way. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.

[192] Chenglong Wang, John Thompson, and Bongshin Lee. Data Formulator: AI-Powered Concept-Driven Visualization Authoring . *IEEE Transactions on Visualization & Computer Graphics*, 30(01):1128–1138, January 2024.

[193] Julie Holliday Wayne, Hoda Vaziri, and Wendy J. Casper. Work-nonwork balance: Development and validation of a global and multidimensional measure. *Journal of Vocational Behavior*, 127:103565, 2021.

[194] Ariane G Wepfer, Tammy D Allen, Rebecca Brauchli, Gregor J Jenny, and Georg F Bauer. Work-life boundaries and well-being: Does work-to-life integration impair well-being through lack of recovery? *Journal of Business and Psychology*, 33(6):727–740, 2018.

[195] Mark Whooley, Bernd Ploderer, and Kathleen Gray. On the integration of self-tracking data amongst quantified self members. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014*, pages 151–160. British Computer Society BCS, 2014.

[196] Kapo Wong, Alan HS Chan, and SC Ngan. The effect of long working hours and overtime on occupational health: a meta-analysis of evidence from 1998 to 2018. *International journal of environmental research and public health*, 16(12):2102, 2019.

[197] Di Yan, Jacky Bourgeois, Yen-Chia Hsu, and Gerd Kortuem. Paircolator: Pair collaboration for sensemaking and reflection on personal data. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.

[198] Kathryn Yorkston, Kurt Johnson, Estelle Klasner, Dagmar Amtmann, Carrie Kuehn, and Brian Dudgeon. Getting the work done: a qualitative study of individuals with multiple sclerosis. *Disability and Rehabilitation*, 25(8):369–379, 2003. PMID: 12745946.

[199] Jun Zhu, Sruzan Lolla, Meeshu Agnihotri, Sahar Asgari Tappeh, Lala Guluzade, Elena Agapie, and Corina Sas. A systematic review and meta-analysis of research on goals for behavior change. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery.