People Like Me: Designing for Reflection on Aggregate Cohort Data in Personal Informatics Systems

CLAYTON FEUSTEL, Georgia Institute of Technology, USA
SHYAMAK AGGARWAL, Georgia Institute of Technology, USA
BONGSHIN LEE, Microsoft Research, USA
LAUREN WILCOX, Georgia Institute of Technology, USA

Increases in data complexity in personal informatics systems require new ways of contextualizing personal data to facilitate meaningful reflection. An emerging approach for providing such context includes augmenting one's personal data with the data of others "like them" to help individuals make sense of their data. However, we do not yet understand how an individual's self-reflection process is affected when the data of others is made available. In this paper, we investigate how people reflect on three types of personal data when presented alongside a large set of aggregated data of multiple cohorts. We conducted personal and cohort data reviews using a subset of participants from a mobile-sensing study that collected physical activity, digital social activity, and perceived stress, from 47 students over three weeks. Participants preferred to use characteristics of the data (e.g., maxima, minima) and graphical presentation (e.g., appearance of trends) along with demographic identities (e.g., age, gender) when relating to cohorts. We further characterize how participants incorporated cohort data into their self-reflection process, and conclude with discussion of the implications for personal informatics systems that leverage the data of "people like me" to enable meaningful reflection.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**;

Additional Key Words and Phrases: Personal Informatics, Self-Reflection, Qualitative Study, Health and Wellness, Cohort Data

ACM Reference Format:
Clayton Feustel, Shyamak Aggarwal, Bongshin Lee, and Lauren Wilcox. 2018. People Like Me: Designing for Reflection on Aggregate Cohort Data in Personal Informatics Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 107 (September 2018), 21 pages. https://doi.org/10.1145/3264917

1 INTRODUCTION

Advances in mobile, wearable, and environmental sensors have opened up new areas of research for ubiquitous computing designed to promote personal well-being. Individuals now have the ability to collect multiple streams of data relating to their behavior [34, 64], mood [22], and physiology [57]. Through purposeful review and reflection of these data streams, people can gain personal insights [43], especially with visual data exploration [11].

Research in personal informatics systems describes the processes that play a vital role in gaining personal insights [11]. While review of data collected over time is an important fundamental capability, being able to see one's data contextualized in the data of others enables people to develop different types of insights [6].

Authors' addresses: Clayton Feustel, Georgia Institute of Technology, North Ave NW, Atlanta, GA, 30332, USA, clayton.feustel@gatech.edu; Shyamak Aggarwal, Georgia Institute of Technology, North Ave NW, Atlanta, GA, 30332, USA, shyamak@gatech.edu; Bongshin Lee, Microsoft Research, 14820 NE 36th Street, Redmond, WA, 98052, USA, bongshin@microsoft.com; Lauren Wilcox, Georgia Institute of Technology, North Ave NW, Atlanta, GA, 30332, USA, wilcox@gatech.edu.

2474-9567/2018/9-ART107 $15.00
https://doi.org/10.1145/3264917

Some personal informatics systems now enable these capabilities by allowing individuals to view not only their own data but also the personal data of others "like them"—opening up opportunities to make comparisons, set goals, and assess progress toward goals relative to a cohort [2, 42]. Indeed, the ability to "see oneself" in a dataset constructed by peers can be a powerful source of personal insight while potentially promoting ongoing engagement with personal informatics systems [35, 36].

Personal self-tracking communities, especially the Quantified Self community, have a rich history of sharing strategies, tips, and data with other trackers [12]. Recent large, community-scale deployments of activity monitoring [56, 65], made possible through the increasing ubiquity of commercial trackers, have created data repositories of unprecedented size. Integrating this large-scale personal tracking data to improve personal reflection is an ongoing challenge in personal informatics research. Some commercial tracking applications offer individuals access to large aggregate datasets, segmenting them by demographic identities. Fitbit, for example, encourages people to compare themselves to others who share the same age and gender to discover new "sleep insights" [26].

Supporting reflection in personal informatics systems goes beyond simply providing access to collected personal or social data [19]. As we describe in Section 2, research into the nature of reflection has suggested a set of techniques and design approaches to scaffold reflection on personal data [7, 28, 61]. Researchers have also shown that careful and intentional design for reflection can create systems that encourage positive reflective practices [1, 27].

To effectively design these systems, it is important to understand how individuals engage and interact with the collected data. For personal informatics systems that incorporate others' data alongside personal data, many of these behaviors have yet to be understood. These gaps in knowledge make it difficult to design systems that account for the potential—both positive and negative—of leveraging this data. In this paper, we begin to bridge this gap by conducting data review sessions, using real-world physical, social, and stress data collected from 47 undergraduate and graduate students, with 10 participants of the data collection study [9]. To characterize the behaviors of individuals interacting with personal and cohort data, we provided the participant's personal data alongside the same types of data of others (who also participated in the data collection study, at the same time). We used time-series visualizations of their personally-collected data and engaged each participant in reflective tasks, asking them to first investigate phenomena that were uniquely interesting to them and then construct a personal goal.

Through study sessions consisting of both think-aloud reviews and a semi-structured interview, we found that participants looked for shared characteristics (e.g., same age group with similar work habits) to relate with a cohort, but that the similarity of the graphical trends in data strongly affected the participant's bond to the group. When participants began data exploration, they first identified days that showed extreme high or low levels of activity or stress and used these moments as "anchors," reasoning about these extremes first to interpret the surrounding data. This strategy extended to how they reflected on cohort data: participants would begin exploring group data by evaluating why certain extremes were in their own data but absent in the cohort data—or vice versa.

The key contributions of this paper are twofold: (1) We describe when and how participants used others' self-tracking data to anchor their own reflective inquiry. We then draw on their behaviors and reflective commentary to characterize how participants chose with whom to identify when presented with the data of multiple cohorts; and (2) We discuss implications for the design of systems that support the reflective behaviors of individuals when interacting with the self-tracking data of others.

2 BACKGROUND AND RELATED WORK

Several streams of ubiquitous computing and HCI research focus on how to promote personal awareness of health and wellness through data collection and review. Personal informatics, design, and social and behavioral sciences all contribute their own understanding of what it means for users to reflect and act on self-tracking data. Since the definition of reflection is somewhat elusive [7, 28], we draw from all these domains to provide a holistic view of reflection. Below, we situate our study with respect to each of these areas.

2.1 Reflection in Personal Informatics

Personal informatics technologies often leverage mobile, ubiquitous sensing while applying theories of behavior and user-centered design to improve individuals' awareness of their health [41, 52, 58] and everyday behaviors (e.g., physical activity [16, 47, 63] and sleep [4, 10, 40]).

In models of personal informatics, reflection was initially treated as one stage in a larger process. In Li et al.'s early model of the use of personal informatics tools, *reflection* precedes *action*, the ultimate stage. Yet, our understanding of the role of reflection in personal informatics systems continues to evolve. Li et al.'s updated model [43] includes two distinct, goal-directed phases inherent in the reflection stage: (1) a *discovery* phase, in which data is evaluated to identify goals and corresponding behaviors to change, and (2) a *maintenance* phase, in which individuals ensure that they sustain a desired state. Choe et al.'s study of the practices of self-trackers found that reflection often occurs concurrently with collection, suggesting a more integrated role of reflection not limited to a specific stage [12].

Building on Rooksby et al.'s efforts to characterize real-world *lived informatics* [60], Epstein et al. constructed an expanded model: a stage-based, non-linear, and iterative model of lived informatics that situates self-tracking in everyday life [21]. Notably, Epstein et al.'s model is not limited by a behavior-change slant. It accounts for individuals motivated by behavior change, as well as those who are instrumenting a particular activity or tracking out of curiosity. In this model, reflection is one of three components that also include collection and integration of data—activities that often occur simultaneously. Together, these components support *tracking and acting* in an ongoing way. However, there is an underlying assumption in the model that individuals are interacting with only their own data. Making the personal data of others visible changes the way users interpret and act on their own data [2, 6]. Knowledge of how these interpretative processes unfold—which we address in this work—can point to useful augmentations to personal informatics models.

2.2 Reflective and Interpersonal Informatics

Bales and Griswold proposed *interpersonal informatics* to address the gap in our understanding of how social data in personal informatics systems could benefit users [2]. Thus far, interpersonal informatics has focused on understanding the technical and social concerns of implementing such systems (e.g., privacy considerations) [2]. Taking these into account, we expand on interpersonal informatics by exploring the mechanisms of reflection that arise when multiple streams of cohort data are made available to users of a personal informatics system.

Reflective informatics is an approach to more broadly understanding the processes that underpin human engagement in self-reflection. It synthesizes research in education, philosophy, cognitive science, and critical design, to suggest a set of dimensions inherent in reflection: *breakdown*, *inquiry*, and *transformation* [5]. Not constrained to a specific domain, reflective informatics can be used to better understand the reflection stage present in all personal informatics models. These dimensions are intentionally broad, however, and have many incarnations depending on the application design (i.e., an application can choose to support breakdown through comparison with historical data or through intentionally changing the interaction with an interface—both create unexpected results which prompt reflection [5]). By allowing users to interact with their personal data as well as the data of others in a more open-ended way, we can explore specific occurrences of these dimensions at work and uncover mechanisms that interact with them.

2.3 Designing for Reflection

Researchers recently began to consider reflection as a primary design goal rather than a means to an end (in which an outcome like behavior change is the measure of success) [7]. Gouvia et al. found that participants using a personal informatics system rarely engaged in deep moments of reflection, instead preferring brief 5-second interactions, *glances*, when using the system [32]. Personal informatics, therefore, is left with the question of how to engage users in moments of deeper reflection. Fortunately, complementary work offers descriptions and frameworks that designers can draw on to encourage reflective behavior.

Fleck and Fitzpatrick characterized reflection as a multi-level framework with specific behaviors and activities that correspond to each level [28]. With this framework, researchers have turned their attention to the *capabilities* needed to support reflection practices in personal informatics systems. Rivera-Pelayo et al. identified three categories of support for focused practices in reflection: *tracking cues*, *triggering*, and *recalling and revisiting experiences* [59]. Cuttone et al. proposed heuristics for visualizations of personal data that support reflection, emphasizing support for providing the means to answer questions about patterns and relationships in the data, and to view the data from multiple perspectives [17]—capabilities found to be important by Fleck and Fitzpatrick [28]. Slovak et al. extended this work on design heuristics by proposing a framework that suggests ways of scaffolding for transformative reflection [61]. These descriptions provide guidelines for developing systems that promote reflection, yet they are broad design recommendations that—intentionally—do not consider the specifics of the domain in which they are applied. Such intricacies are necessary to fully support users in their reflective activities, and to understand the implications—both positive and negative—of leveraging the data of others to support awareness and understanding of one's own data.

2.4 Behavior Change through Social Data

Much of the prior work in designing interactive systems for health and wellness leverages tracking of personal data with feedback or data review capabilities. A significant amount of prior work also incorporates personal data of others. However, there are distinct differences in the approaches applications take to make the "personal data of others" visible. Often, an underlying intent inherent in the design shapes choices about how the data of others is incorporated. For example, many systems will use others' data to promote behavior change through social comparison [25], explicitly inviting competition between individuals [44] or social support [15]. Many of these systems expose personal identities in order to facilitate competition or social support, but more recent work has applied these concepts to applications where personal identities are protected through pseudonyms and avatars [29], finding that this social data also brings about behavior change. Miller's synthesis of theories of social behavior suggests that peer support is vital in changing health attitudes and behaviors [46]. Data sharing and social awareness are often incorporated into applications specifically to leverage the *behavioral influence* of peers and role models [45]. One strategy used to achieve this influence includes tracking and sharing of one's performance data (e.g., physical activity levels) to promote behavior change [25]. Some applications in this vein embed behavioral goals in team-based games, explicitly inviting competition between users [44]. Another strategy integrates peer communication and personal data transparency (but without competition-driven incentives) [30, 47, 63]. This strategy can foster social cognitive reasoning [3] and social support [15]. This form of behavioral influence can emerge organically in online social settings, when users actively share health data (e.g., broadcasting their fitness activity) [50].

Ploderer et al. recognized the strategies outlined above, distilling five approaches to promoting social interaction and reflection in behavior-change applications [53]. In their framework, *social traces* refer to data depicting patterns of other users of a computing application or system. While the traces rely on input from a larger group, they primarily benefit the individuals comparing themselves with others. The authors contrast this with *social support*, referring to various forms of deliberate exchange between people aimed at enhancing their well-being.

Our work expands on these theories to explore how individuals incorporate anonymized aggregate data into their self-reflection process. Rather than focusing on how this social data can be leveraged to encourage behavior change, our interest is in exploring how social influence from others' data informs the underlying processes used by individuals when they reflect on their personal data.

2.5 Promoting Personal and Social Awareness

Thus far, the social applications of data we have touched on—including those utilizing a large set of aggregate data as well as pre-defined groups, and with either anonymized or identified data of others—all enable social awareness to implicitly promote behavior change. However, emerging work on social data awareness and reflection makes an important distinction between more prescriptive designs for social data reflection versus more "open-ended reflection" [6, 48, 53].

Researchers are now also beginning to explore many ways that "personal" or "self-tracking" data mediates social life [20] and, in parallel, how social relationships can influence reflection on personal data. Graham et al. proposed an open-ended reflection through "peer review" of self-tracking data. In a study with 15 people who were free to choose what type of data to collect, and how to log it, they found that peer review, analysis, and feedback led to *shared reflection*, which motivated participants to keep up with their data collection while providing sources of surprise, advice, and moral support [33].

In this paper, we primarily treat *reflection* as a process of considering one's present, past, or planned wellness-related behaviors, attitudes, and experiences [51], focusing on designing for reflection on collected data [1]. Using this lens, we investigate how people integrate the personal data of others alongside their own data during retrospective reflection, highlighting unique activities they engaged in and implications for design.

3 STUDY

We conducted personal and cohort data reviews using the data generated from a mobile-sensing study, which collected 15 types of data from 47 undergraduate and graduate students for three weeks, through their mobile phones [9]. Our goal in this study was to gain a preliminary, qualitative understanding of the ways in which the data of others shape an individual's reflective process when seeking insights about their data or setting a goal. Thus, we decided to observe people in a private, lab-based setting as they explored their own data and to give them time to make reflective comments to researchers. University students represented an important group for our study: these participants are in a transitional period during which they shift into more autonomous managers of their daily routines and lifestyle choices. This period is characterized by a deepening sense of personal identity, making the exploration of "people like me" for this group compelling. At the same time, they share lifestyle characteristics that make meaningful cohort exploration possible (e.g., socializing in similar places, sharing an academic calendar). Our study participants performed individual retrospective reviews with *their own data* complemented with the same types of data from others who participated in the mobile-sensing study.

3.1 Personal Data from the Mobile-Sensing Study

The mobile-sensing study collected trait survey data on *demographics*, *emotion regulation*, *social media activity*, and *resilience* [9]. In addition, it collected "passive" and "active" data through students' mobile phones. The "active" data included responses of periodic Ecological Momentary Assessments (EMA) through a commercial platform (Quedget [55]) to prompt participants to log self-reported data on *mood*, *perceived stress*, and *current activity* using lock screen interfaces. The mobile-sensing study participants contributed 6,080 self-reported responses over the three-week period (avg response rate of 11.47 measures/day).

The passive sensing leveraged sensors on the phone, using a mobile instrumentation research framework (AWARE [24]) to capture *accelerometry, GPS, ambient noise levels, call and message meta-data, screen state,* *application history, and physical activity* (provided through Google's DetectedActivity API [31]). The passive data streams were automatically collected at predetermined polling rates from the participants' phones throughout the duration of the study. Participants did not interact with their "passive" or "active" data while participating in the mobile-sensing data collection. We were thus able to capture, in the study sessions we describe here, how the participants responded to and made sense of their personal data when shown to them for the first time.

3.2 Study Materials

To perform retrospective data review sessions, we constructed three measures of physical and psychological wellness: *daily physical activity*, *digital social activity*, and *perceived stress* from the mobile-sensing data. *Daily physical activity* describes the number of minutes a day an individual has been active by adding together system-detected activities of on-bicycle, on-foot, running, and walking. *Digital social activity* examines digital sociability by evaluating the call and message meta-data for the number of minutes on the phone and the number of text messages sent and received each day. *Perceived stress* indicates daily stress on a scale of 1–5, determined through PSM (Psychological Stress Measure) [14], delivered through EMA prompts. We selected these measures based on their presence in design research in the health and wellness domain. We also wanted to understand interaction differences that arise due to changes in the type of data presented: phenomena of current relevance to personal informatics [21, 39, 60].

To construct aggregate cohort data, we created seven groups to segment the full mobile sensing participant pool: Individual, Everyone, Age, Gender, Resilience, Social Media Use, and Perceived Stress. We further delineated these groups as outlined below, into Individual (their personal data) and 15 cohorts (Table 1). Everyone (all participants in the mobile sensing study), Age, and Gender are demographic groupings commonly used in both research and commercial applications. Resilience, Social Media Use, and Perceived Stress, on the other hand, present new lifestyle and wellness categories for participants to explore. All of these three non-standard groups come from evaluation of a comprehensive pre-test conducted prior to participants joining the mobile-sensing study that covered a wide range of psychological and interpersonal factors. Each participant was assigned to a cohort in the Perceived Stress group (based on their DASS21 pre-test score) and also had regularly-collected EMA data on *perceived stress* that was included as a data type for review in our sessions.

We acknowledge that there is no single correct way to stratify a population based on grit, stress, or social media use. To enable exploration of a number of relevant cohorts, we created abstractions based on distribution statistics from the mobile-sensing study population [9] and grouping criteria found from prior psychometric studies [8, 18, 38] to make it easier for participants to understand and interact with these data (details in Table 1). We are primarily interested in understanding the potential for psychographics and social cohorts to provide meaningful ways to assess similarity with a group that are not afforded by demographic categories or pre-determined networks alone. The qualifications we used are not exclusive and we see many alternative ways for researchers to represent these non-traditional cohorts.

3.3 Participants and Study Prototype

Following IRB approval of the study, we recruited 10 participants (four females) for individual personal and cohort data review sessions (reaching data saturation) from the 47 mobile-sensing study participants. Participants were required to have a minimum of seven days' worth of physical activity and PSM data as well as call and message meta-data. Participants ranged from 18–25 (*mean* = 21) years of age and included seven undergraduate students and three graduate students.

Participants visually explored their own data with the option of reviewing the data of 15 additional cohorts (Table 1) through time-series visualizations we created using Tableau [62]. We chose to display the data as a time series, because of the significance of this representation in personal informatics systems. In addition to

Table 1. Seven groups used to construct cohorts, comprising 15 possible cohorts in addition to Individual data. We used distribution statistics from the mobile sensing study population [9] to stratify the Age group into three cohorts. Social Media Use is stratified based on results of a pre-study questionnaire containing 10 Likert-style (1–5) questions [9]. We used grouping criteria derived from prior psychometric studies, cited in the table, to stratify Resilience and Perceived Stress groups into Low, Medium, or High cohorts.

Group	**Measure**	**Cohort**
Individual	—	—
Everyone	—	Everyone
Age	Pre-Study Questionnaire	19 & Younger | 20–24 | 25 & Older
Gender	Pre-Study Questionnaire	Male | Female
Resilience	Grit-S [18]	Low | Medium | High [18]
Perceived Stress	DASS21 [37]	Low | Medium | High [8, 38]
Social Media Use	Pre-study Questionnaire	Low | Medium | High

being the most prominent form of visualization used by personal tracking systems [49], time series graphs are a visualization type that provide users the ability to observe patterns over time—a critical ability for assisting documentary trackers in discovering relationships between data types and to help goal-driven trackers monitor progress towards a goal—without steeping the user in complex statistical or graphical representations that may exceed the average user's graphical literacy [23].

The study prototype (Figure 1) enabled participants to add and remove data on *daily physical activity*, *digital social activity*, or *perceived stress*. Participants can show or hide all cohorts as well as their own data (Individual data). For example, participants can choose to explore their *daily physical activity* for a nine-day period by selecting Individual. They can choose to compare these data to the average daily physical activity data of those in a Low Stress, Medium Stress, or High Stress cohort (or any combination) during the same nine-day period. The cohort in which a participant is a member is indicated in the selection interface, but all cohorts are available for review. Participants can also remove Individual data, reviewing only the data of other cohorts they selected.

When a cohort is selected, a line graph showing the corresponding data appears in the main view (Figure 1c) and a row is added to the Selected Categories section (Figure 1e). Participants can also hover over a data point for more details, click on a line to highlight it, or click on a cohort in the Selected Categories section to bring the corresponding graph to the foreground. The interface for *digital social activity* is similar, except it has two graphs, one for calls and one for text messages, as shown in Figure 1b. The graphs representing *digital social activity* function independently, thus allowing participants to add different cohorts for *calls* and *messages*.

3.4 Study Design and Procedure

The study consisted of three parts covering the three wellness measures: *daily physical activity*, *digital social activity*, and *perceived stress*. Each participant was seated in a private office with one researcher, who sat diagonally across from them and had a peripheral view of the screen (to answer questions and ask for clarifications to think-aloud comments).

Participants were first given an overview of the prototype by the researcher (how to add data about themselves or others to the primary viewing window, how to get more detail for a data point, etc.) and then provided an opportunity to explore the functionality of the prototype and ask questions, without being recorded. Questions asked by the participants included how to perform certain functions and clarifications about how cohorts were formed (e.g., how their resilience cohort was determined). When the participant was ready, the first wellness measure, *daily physical activity*, was loaded into the interface. Each part involved a think-aloud session and a semi-structured interview. For each measure, participants were shown their data and asked to perform two

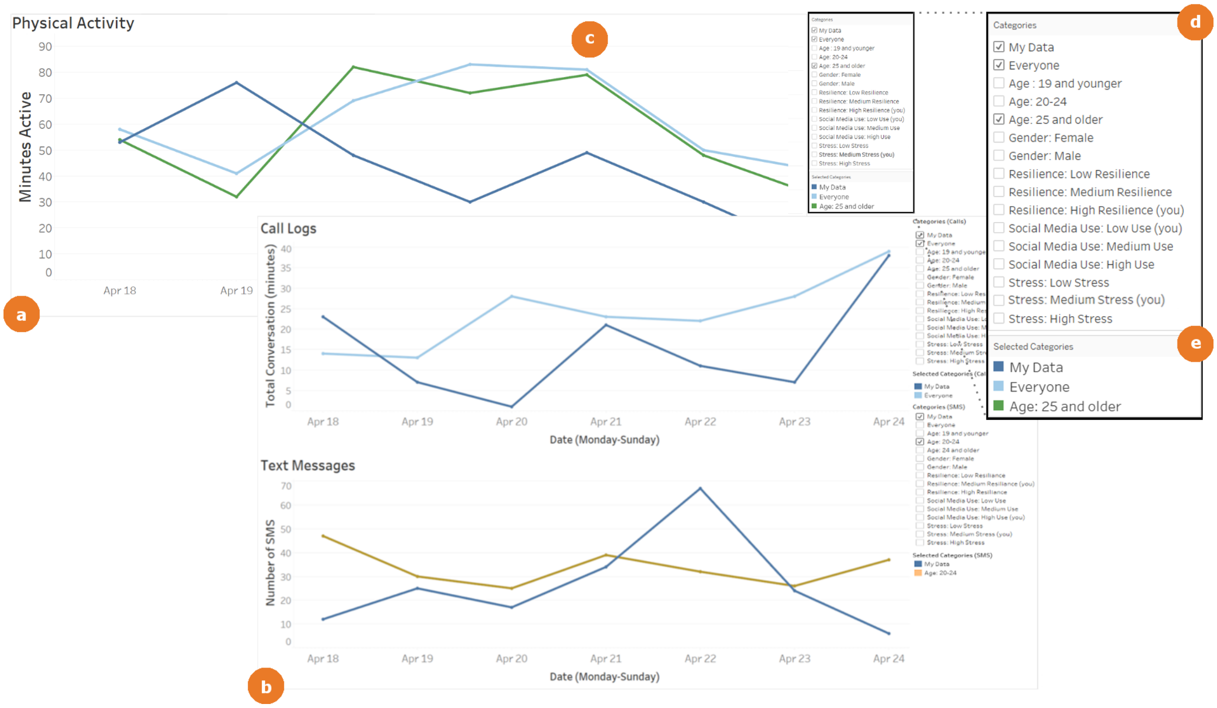

Fig. 1. Interface for displaying (a) *Daily Physical Activity* and *Perceived Stress*, (b) *Digital Social Activity*. User can interact with the visualization by viewing data (c), selecting cohorts (d), and highlighting active cohorts (e).

independent tasks using the prototype user interface—(1) Exploration: investigate any interesting phenomena in their data, and (2) Setting a goal: after the participant indicated that they were finished, set and specify a goal for the near-term using the data (e.g., a participant looking at *daily physical activity* may say that she wants to be active 10 more minutes a day). Participants could take as long as they choose for the Exploration task and after they had indicated completion of it, the researcher introduced the Setting a goal task.

We chose these tasks based on Epstein et al.'s findings that one's motivations for self-tracking (e.g., behavior-change oriented versus curiosity-driven) change the way individuals interact with their tracking data [21]. An individual's motivations are not static, however, and participants may shift from one motivation to another. We accommodated the mindset of both a curiosity-driven tracker (by asking about any phenomena of interest first) and a behavior-change-oriented tracker (by asking participants to set a goal).

At the end of each task, participants were asked a set of questions to elucidate their spoken responses and interactions. Participants were given a $20 Amazon gift card for their time. The session was audio-recorded and all participant interactions with the application were captured through screen-recording software. The session lasted 30–60 minutes (*mean* = 43.38).

4 FINDINGS

Interviews were transcribed verbatim and screen capture data was time-stamped for cross-reference during qualitative analysis. Two researchers coded all transcripts in a bottom-up, iterative fashion: they discussed emergent codes, resolving disagreement and refining codes in-person. Codes were then cooperatively grouped by both researchers into higher-level clusters. Labels describing the groupings were then constructed to determine higher-level themes once consensus was reached. Seven themes comprising three categories emerged from our qualitative analysis (Table 2). In Table 2, descriptive counts are provided with themes to supplement qualitative data. However, these counts are not absolute as responses were undirected (we can only count participants who made think-aloud responses related to the theme during our interview session).

Table 2. The seven themes and three categories that emerged from qualitative analysis of participants interacting with aggregate social data. Categories provide the high-level type of activity that participants engaged in, while themes describe specific interactions or interests articulated by participants. The Context column shows the data type the participant is looking at: *daily physical activity*, *digital social activity*, or *perceived stress*) and what task was being performed (observation or goal-setting). Brackets included in the Theme column contain the number of participants who provided undirected think-aloud explanations that match the theme.

Category	**Theme**	**Example Quote**	**Context**
Cohort Selection	Selection is based on shared personal or lifestyle characteristics [7 of 10]	*"... being able to compare it to both males and 20–24 sort of gives me an idea where I am compared to someone with a similar build to me."* (P7)	View: Daily Physical Activity Task: Exploration
	Searches for graphs that have a similar pattern to personal data [8 of 10]	*"So I've got a line for mine, right? And what I'm trying to do is compare that trend line with different groups to see which trend line is my trend line most similar to."* (P6)	View: Daily Physical Activity Task: Exploration
Identifying Insights	Interested in patterns between cohorts [5 of 10]	*"Those with high stress seem to have more erratic [patterns], it seems to be more erratic in terms of their activities whereas those with low and medium seem to be similar. Their highs are not that high, their lows are not that low"* (P6)	View: Daily Physical Activity Task: Exploration
	Interested in relationship between cohort and data type [6 of 10]	*"For physical activity, for example, you don't really care about psychological resilience. Some other person might care. But stress is, I think, is more of a mental concept, so I would look at mental factors."* (P4)	View: Perceived Stress Task: Exploration
	Drawn to extremes in data (peaks and valleys) [7 of 10]	*"So, I'm not really sure why I'm particularly high stress, what's the 24th and 25th? This was the last days of classes ..."* (P7)	View: Perceived Stress Task: Exploration
Goal Setting	Goal based on average or "normal" data [5 of 10]	*"I like to have categories for social media usage, specifically, just to have some boundaries for what is considered typical behavior ... to compare to my own activities."* (P5)	View: Digital Social Activity Task: Setting a goal
	Goal based on pre-existing personal desires and external constraints [6 of 10]	*"And my goal is to really, instead of texting them, actually find time to actually engage with them on top of the conversation, face to face instead of just through messages ..."* (P3)	View: Digital Social Activity Task: Setting a goal

4.1 Cohort Selection

Participant interaction in selecting and deselecting cohorts revealed how individuals formed relationships with aggregate cohorts, and how these relationships shaped the perceived utility of cohort data in the self-reflection process. As expected, demographic identities served as a strong characteristic that drew many participants to include cohorts in their reflection. Participants were also driven by what the data showed. Similarity in characteristics between the individual's data and others' (described in detail below) as well as the *graphical pattern* of the data influenced whether participants related to a cohort and subsequently, whether they relied on that cohort data during the self-reflection process.

4.1.1 Defaulting to Personal Characteristics. Participants initially selected cohorts based on how similar they perceived the *individuals* comprising the cohort to be to themselves. The factors for determining similarity differed between participants, but all were centered around whether others shared personal characteristics with them. P1, for example, explained how her notions of strength differences between males and females prompted her to include the Female cohort in her initial exploration of the *daily physical activity* data: *"[I am] comparing [my] stats with Females ... because I feel like guys are more active than girls, just 'cause they're stronger, they have more stamina."* After briefly adding the Male cohort to the graph, however, P1 noted that the data did not reflect her expectation: *"But the data is showing a little bit ... oh. I'll be honest, I would expect the Males to be way higher up on this chart ... [but] it's much lower than Females."*

In spite of this, when P1 was asked to set a goal she once again expressed interest in the Female cohort, reiterating a similar sentiment about gender differences in strength: *"For physical activity, I would select Female because males are stronger than females and I can't compare with a male because we have different bodies ... different stamina, different strengths."*

Participants used Age cohorts to find others with similar living and work habits. P2 explained why he did not add the 19 & Younger age cohort for any of the three data types, *"[I] just assumed that they may have a different lifestyle."* P1 echoed P4's concern for lifestyle differences by selecting the 19 & Younger cohort because age played a role in how much physical activity she expected to see: *"Freshmen who live in the dorm, they walk a lot ... age 20 to 24 and age 25 and older, they have less [activity]. They probably commute to school by car."*

4.1.2 Searching for Similar Graphical Patterns. In addition to shared personal characteristics such as demographics, participants searched for cohorts that displayed similar graphical patterns in their data. P6 explained her process of attempting to identify cohorts that had similar physical activity habits (shown in Figure 2): *"So I've got a line for mine, right? And what I'm trying to do is [to] compare that trend line with different groups to see which trend line is my trend line most similar to."*

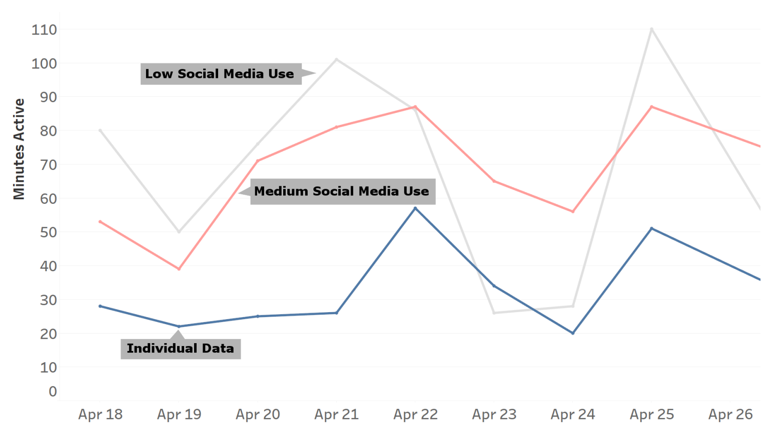

Fig. 2. View of P6 highlighting Medium Social Media Use cohort (red line), which she identified as similar to her own data (blue line). Low Social Media Use, P6's placed cohort (gray line) was stated to be "highly erratic" compared to her data.

When investigating her *perceived stress* data, P4 initially identified the Medium Stress cohort as a "guideline," but quickly changed her mind because the graph did not reflect the same extremes present in her own data: *"I would like to see Stress categories, but I think medium is a good guideline ... I guess I would go with, like, Medium Resilience. ... Actually, no. I would want to have spikes in that graph, because I know that's more accurate."*

Participants additionally used data similarity to confirm assignment to a cohort, particularly for the non-demographic groups of Social Media Use, Stress, and Resilience. P9 went through a complex process in which her use of the High Resilience cohort fluctuated between data types due to changes in graphical similarity between data types. When reviewing her *digital social activity*, P9 noted that her High Resilience cohort demonstrated dissimilar texting behavior from herself. This single difference impacted how P9 related to the cohort as a whole, causing her to forego including High Resilience as useful when reflecting on *digital social activity*:

P9: *"Medium resilience likes to text more and high resilience likes to call more. That doesn't fit me, if I'm high resilience."*

Researcher (R): *"Mm-hmm. Does that kind of affect how strongly you relate with this high resilience?"*

P9: *"Probably ... I feel like maybe it's not a great indicator for at least predicting my communications with others."*

When P9 evaluated her *perceived stress data*, however, she noted significantly more aligned data points: *"High Resilience is going up and down, which is kind of like me, so maybe I do fit this High Resilience more here than in my call log data."* Accordingly P9 changed how strongly she identified with the cohort, deciding, in the end, that High Resilience was an important category for evaluating *perceived stress*: *"... Maybe I was surprised that I was put in High Resilience. But seeing that I am kind of, like, the same general pattern here—given my limited data—it's kind of like, 'Okay, well, maybe I am more High Resilience versus the other two.' "*

4.2 Identifying Insights

In evaluating their data, participants performed exploration tasks that were unique to systems that incorporate aggregate social data. They looked for similarity between patterns in their data and patterns in the the data of the cohort to which they belonged. They also looked for similarity between patterns in their data and the data of other cohorts. Both of these activities played an integral role in cohort selection. Also integral to exploration was the identification of data points that were perceived as outliers and that contradicted a participant's expected results. We found that these provided "anchors" for further exploration of personal data and guided initial analysis of others' data.

4.2.1 Seeking Patterns between Cohorts. Participants were interested in uncovering relationships between cohorts, even those that they were not a part of. Many of these observations were not immediately used in comparison to the participant's data, but rather general observations about how cohorts compared to one another. P9 looked for relationships between cohorts and proposed explanations for the patterns. When looking at *daily physical activity*, she noticed a relationship between resilience cohorts remarking, *"It looks like the more resilient you are, the less active you are. Maybe people who are less resilient need to walk around, or like, workout more or something."* In addition, P9 noted that *"The more social media you use, the less active you are ... you're not really tweeting about your hikes."*

P6 shared a desire to find a relationship between cohorts when viewing *daily physical activity*, evaluating the behavior of all possible cohorts in the Perceived Stress group: *"Those with High Stress seem to be more erratic in terms of their activities whereas those with low and medium seem to be similar. Their highs are not that high, their lows are not that low."* She did not immediately know how she would utilize this additional insight explaining that she *"was just trying to compare all three groups"* even if she did not yet know *"how to digest this information."*

4.2.2 Relating Cohorts and Data Types. In addition to examining relationships between cohorts, participants were interested in understanding the relationships between the data type and the cohorts. By and large participants sought to assess correlation between cohort and data type as a means for determining how relevant a group was for their self-reflection. Occasionally, these relationships were determined by personal beliefs, while at other times participants used the data itself to decide if there was a connection between cohort and data type.

When explaining why P4 used the High Resilience cohort for setting a stress goal, but did not use any Resilience cohort for *daily physical activity*, she highlighted the significance of the cohort's relationship with the data type: *"For physical activity you don't really care ... [about resilience] But stress is more a mental concept, so I would look at mental factors."* The Age cohort played a significant role in P1's physical activity exploration and goal setting—with P1 opting to use her age group and gender as guidelines for her *daily physical activity* goal—yet she did not want to use any of the Age cohorts for exploration during review of *digital social activity*, because *"it doesn't really matter your age ... Everyone should get in touch with their friends and family."*

If participants did not have a preconceived notion of how a cohort related to a data type, they would seek out this information through exploration of the data. Looking at *daily physical activity*, P7 investigated the relationship between resilience and physical activity by adding all three resilience levels (low, medium, and high) and looked for patterns. He did not uncover any discernible relationship, explaining *"I don't really see much of a correlation ... they all seem roughly average."* At the end of physical activity exploration, P7 described the resilience cohorts as "arbitrary" and stated his preference for age where he could easily imagine a relationship in which *"Someone who's older ... [has] less time to be active."*

4.2.3 Anchoring Exploration through Extremes in Data. Participants consistently used extremes in the data as a way of bootstrapping exploration. P1 described how she utilized peaks in her physical activity to begin exploration: *"So, I'm gonna first click on, my data ... I'm trying to analyze what this high point is."* After receiving his *daily physical activity* graph, P2 was immediately drawn to the highest activity in his data and sought deeper explanations of these data points before proceeding to more findings: *"I tend to walk a lot on days that I'm traveling. So I think I actually had a flight, somewhere at the end of this week, maybe here. But I'd be interested in knowing what these peaks are as to what's going on there."*

Using extremes to anchor the exploration of health data extended to participant interaction with cohorts. P7, for example, began exploring his *daily physical activity* data by noticing that his peak physical activity occurred on a day of relatively low physical activity for others (Figure 3): *"I have a spike here, most have the exact opposite*

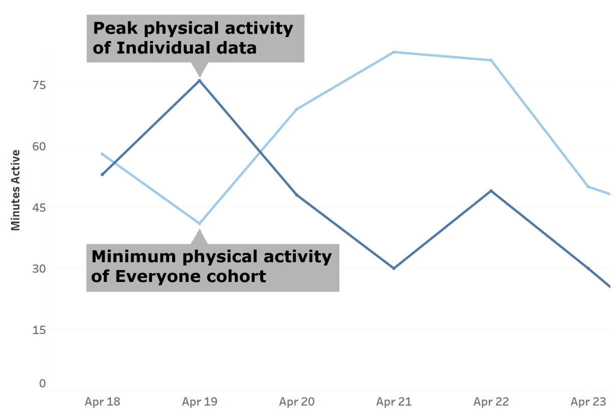

Fig. 3. P7 exploring Individual data (dark blue) compared to Everyone (light blue). He begins discussing data captured on April 19th, pointing out discrepancies between a personal high point for him relative to a low point for Everyone. Annotations have been added for clarity.

direction. There's a dip at the beginning ... So everyone goes down, whereas mine goes up." In this case, the difference between his data and what was typical for his chosen cohort drew him into further examination and reflection.

P9 also explored his data by identifying differences in his personal data peaks with those of the Everyone cohort: *"So in terms of everyone, I seem to be like a lot more on the extremes. When I was really active, everyone was kind of active, and when I was really not active, everyone was kind of not active ..."* P6 began exploring her data by first looking at days with the highest perceived stress, providing explanations for each spike. After looking through Individual data, P6 added the Everyone cohort and began looking for matches in "high stress" days: *"[I] seemed really stressed on 25th. Is that when I had the exam? [The] 25th was Monday. Yeah and I had the exam in the afternoon if I remember it right ... And there was some crazy deadlines so that explains the graph. Comparing it with everyone [User clicks Everyone, adding cohort to the display], why is everyone so chill?"*

Quickly losing interest in the Everyone cohort, P6 replaced it with 25 & Older and began exploring the new age-related cohort with the same peak day of perceived stress: *"I'm excited by that [data] point. Everybody was as stressed as I was on that day. In a very weird way that's cool to know that everybody in my age group, probably grad students, had something going on for that point in time."*

4.3 Goal Setting

When participants searched for unique patterns in their data, they often interacted with cohort data to aid them. During the goal setting portion of our study, we observed unique interactions between cohort data and Individual data. Cohort data was used to establish a "global average" to define the bounds of what behavior was "normal" for cohorts that participants already identified with. However, after consulting what was "normal" for these cohorts, they engaged in a complex reasoning process, articulating a number of different factors and values to further interpret and weigh the importance of the cohort's "norm" to form a personal goal. Below, we further discuss the ways that participants used data to formulate their personal goals.

4.3.1 Determining What's "Normal" for Me Based on What's Average for Others Like Me. Participants often used cohort data to determine a "normal" range for their data relative to their chosen cohorts (e.g., using an average of the data of one or more aggregate cohorts to determine normality). When asked to provide a quantitative value for their goal, P1 mentioned: *"... [I] try to reach the average of the two ... based off of Gender and Age."*

Similarly, P5 expressed that he wanted to use the Social Media Use group to help formulate a sociability goal by having *"some kind of boundary for what is typical behavior. To have that to compare with my own activities."* P3 set his goal based on the desire to stay within an acceptable range of others:

P3: *"Oh, okay. On this graph ... I think I did pretty well. So as long as I'm about a '2' or below, out of '4' then uh that's really my goal."*

R: *"Okay and uh why 2?"*

P3: *"Because it's about average. And if I'm about medium on stress, I'm okay with that ... I just want to make sure that I don't go above that."*

The desire to have health activity similar to what others are doing is exemplified by P1, who remarked, *"'Cause I have to be normal compared to others. I don't want to be way off the chart."* In this case, P1 valued "fitting in" with others that she saw as being like her by attaining similar numbers.

4.3.2 Negotiating Data, Personal Values, and Known Constraints. We found that, ultimately, participants weighed their existing motivations, prior knowledge, and external constraints against *daily physical activity*, *digital social activity*, and *perceived stress* data shown by the system, to determine a personal goal that best fit both their historical data and individual needs.

Participants drew on pre-existing personal desires and external constraints during the goal setting task, and we observed an intricate interplay between personal values and data-driven insights. We found that participants filtered discoveries—occurring from review of the data itself—through the lens of personal beliefs about healthy behavior. This often prompted goals based on variables not collected or exposed by the visualization. When participants used external constraints, they did so to frame and modify the preliminary goals they formed, by evaluating its feasibility and refining it as needed.

Participants also explored opportunities to tie their values to their behaviors. When setting a *digital social activity* goal, P3 articulated the importance of co-located communication, framing his goal through this desire: *"My goal is to really, instead of texting them ... find time to actually engage with them face-to-face instead of just through messages. To post a lot of information–that's not really the way to go. So if you want to be detailed and help better explain ideas or work to people, then my goal is to decrease the number of text messages and meet up with the people instead."*

Finally, external constraints played a significant role in how participants used data in setting goals. P6, for example, started her goal setting process by identifying days that she believed to be more accommodating for physical activity and focused initial exploration on these days: *"I'll start with Saturday and Sunday first because my Monday and Friday are driven by where I work ... so I can't control that much."* P7 expressed a similar desire to accommodate his schedule, forming goals based on *"days where I don't walk as much to class."* After having chosen a rough *daily physical activity* goal, P4 used knowledge of her class locations to refine the limits of what she could realistically achieve: *"If you count walking to classes that will probably be a bit more, so yeah, I guess I could easily reach 70, 80 at least."*

Many participants used the average of their selected cohorts as a starting point for setting a goal. However, in some cases, participants placed much more importance on their own data than on the data of others. For example, in speaking about her stress goal, P6 described why she based it solely on the maximum value of her own stress:

P6: *"My highest seems to be three."*
R: *"Mm-hmm."*
P6: *"Right and I don't want to go over that so I want something that's lower than three. Very simple ... Maybe others have higher but I don't care about that because I'm trying to improve mine."*

Similarly, P2 simply placed higher importance on Individual data when compared to cohort data: *"I guess I'm less interested in knowing what other people's trends are and more interested in my own historical trends."*

5 DISCUSSION

Our paper details findings on how individuals reflect on personal data when the data of others, provided as several anonymous aggregate groups, is presented alongside their own. Specifically, we have detailed how individuals select the anonymous groups that they use in their reflection, what processes individuals engage in when identifying insights from the data, and how individuals use the data of others to evaluate their personal data. In this section, we reflect on the implications of these findings and situate our work in relation to the existing work on personal informatics and design for reflection.

Epstein et al. note in their Lived Informatics model that personal tracking is a social activity and many self-trackers already engage with an element of social awareness in their daily use of a personal tracker [21]. Presenting others' data through anonymous aggregate groups is unique from the approaches used by many existing self-tracking systems, which present the data of family members, friends, or individuals whose identities are made visible through personal accounts (e.g., Instagram account page). Promoting social awareness in this manner raises concerns regarding privacy [2] and sharing considerations (e.g., self-censorship [54]). Presenting data through anonymous aggregate cohorts, on the other hand, has allowed companies such as Fitbit to address these concerns while still leveraging the benefits of promoting awareness of others' data. Still, we do not yet understand how, in the absence of pre-existing social ties, individuals identify with anonymous aggregate groups, nor how this anonymity impacts their interaction with others' data when reflecting on their own data. Our study offers preliminary insights that address these questions.

5.1 Relating with Cohorts

When participants first began exploring their data, they were drawn to the demographic groupings of gender and age. Participants saw these demographic groups as a way of identifying others with common lifestyles and personal characteristics, an important factor in finding data that was comparable to their own. As P7 expressed when speaking about his use of the Male and Age 20–24 cohorts, *"[these selections] sort of give me an idea of where I am compared to someone with a similar [physical] build as me."* However, participants' think-aloud and interview responses suggested that seeking out others with similar demographics was only the first step in identifying with a cohort. Participants actively sought cohorts that shared patterns in the data that were similar to theirs, using the cohorts with the most similar visual pattern to reflect on their personal health data.

In a study describing the behavior of individuals who use Instagram to track eating behavior, Chung et al. found that participants were motivated both by people who shared personal characteristics (e.g., student vs. a stay-at-home mom) and by people who were dissimilar in their personal lives, if they shared a health goal [13]. Similarly, our findings indicate that even without knowing the anonymous individuals contributing data, participants are motivated to use tracked data of those with a "similar build" as well as those that share characteristics of their data (e.g., similar peaks and troughs) and visual appearance of trends.

An important facet of these data-driven relationships to a cohort is that the strength of the connection is dynamic, changing across time (as their data changes relative to others' data) and across different data types. This means that providing access to data from a single static cohort is not sufficient for fully supporting individuals in their usage of others' data. The insufficiency of providing access to a single cohort across all data types is further corroborated by the study participants' desire to see cohorts that had a perceived relationship with the data type (e.g., P4 deciding to use resilience in the stress category because *"stress is more a mental concept"*). In order to fully support individuals in reflecting on their personal health, system designers will need to support the dynamic ways that individuals relate to anonymous cohorts. One simple way that system designers can accommodate this need is to provide a plurality of cohort options (as we provided for our participants) to allow them to select the cohorts that they most closely identify with for a given data type.

However, there is a tension inherent in providing individuals access to multiple cohorts. The participants' tendency to select groups that appeared similar to their own implies that participants with low levels of physical activity, for example, may select groups with similarly low levels of physical activity. This is not necessarily a behavior that should be avoided (the effects of this will be discussed in more detail in the following subsection), however, it is one that must be taken into account. With knowledge of this tendency, personal informatics systems can be designed to either mitigate concerns regarding visual similarity, or be mindful when faced with opportunities to exploit visual similarity in order to encourage individuals to more strongly identify with a cohort (e.g., highlighting peaks or troughs that an individual shares with the group).

5.2 Defining What is "Normal"

Our findings indicate that one important use of others' data is to determine what behavior is "normal." Participants commonly used the average of selected cohorts to discover an acceptable range and compared their own data to this constructed notion of normal behaviors and stress. Work that has investigated how individuals incorporate others' data into the reflection process echoes this finding that individuals use these data to determine what is normal. Puussaaret et al. saw that participants using their social platform for sharing and engaging with self-tracked data, Citizens Makers, utilized the data from others to establish a baseline "for normative comparison of health and exercise" [54]. By allowing open-ended access to others' data, Baumer et al. found that participants were able to construct their own notion of "healthiness" through observations of others' behavior [6].

This process of constructing a normal range of health data was an important function of reflection for our participants, too. The average of the selected cohorts served as an important first step in the formation of a participant's personal goal. A range of "typical" data allowed them to see what goal was possible (e.g., P1 setting her physical activity goal to 90 minutes, the average of *Everyone*, even though 60 minutes was her highest recorded day) and to set an upper and lower bound on acceptable ranges (e.g., P3 setting a stress goal of "2 or below" to avoid exceed the average stress level). In fact, this awareness to norms is described by Poderer et al. as one of the defining characteristics of how socially-contributed data can encourage behavior change [53].

However, completely open-ended access to others' data, presented without scaffolding that is grounded in professional expertise and sound research, can lead to misalignment between what an individual considers healthy and what experts consider healthy [6]. Given that participants in our study selected groups that demonstrated similar patterns in the data as their own, there is the potential to use the data of others to confirm that one's own behavioral patterns are "okay" no matter what they are, as long as others share them too. Personal informatics systems must navigate this tension between supporting flexibility in defining one's own behavioral norms while being mindful of the potential to focus on data that promotes the "status quo" for an individual—even if that status quo is unhealthy.

5.3 Bootstrapping Insights and Triggering Reflection

When participants first began exploring their data, they started by identifying days with peaks and valleys. They used these days of extreme data as initial points of reference for further exploration of their own data. This behavior extended to a participant's exploration of others' data as well. Participants would initiate comparison between cohort data and their own in days when extreme differences were apparent. P9, for example, began his comparison to the Everyone cohort remarking, *"when I was really active, everyone was kind of active, and when I was really not active, everyone was kind of not active."*

Viewing participants' behavior through Baumer's dimensions of self-reflection [5], we can see that breakdown events [5] manifested when the participant observed a day of extreme "lows" or extreme "highs." These events drew participants deeper into inspecting the data, and they proceeded to engage in inquiry [5], generating and evaluating possible explanations for the data.

Revisiting Rivera-Pelayo et al.'s framework, these breakdowns serve as *triggers*, prompting the start of a reflective process. However, our findings suggest extensions to these frameworks. In Rivera-Pelayo's framework, *triggers* support reflection, which, in turn, is sustained through *revisiting and recalling personal experiences.* Indeed, revisitation of experience is foundational to reflection [28]. Our study found preliminary evidence that by providing aggregations of the data of multiple people tracking similar data, we gain new perspectives on what *experience* means.

Exploring the abstracted data of others can allow for *social discovery* of common experiences, as part of the reflection process (i.e., when patterns in data are shared between an individual and a cohort). Individual reflection is thus aided by social discovery—which is also its own experience. This social discovery is another mechanism by which reflection can be sustained. Such discovery can also reinforce the revisitation of other personal experiences, thus continuing the reflective process.

Designers of personal informatics systems could use knowledge of such breakdown events and triggers—arising in our study as unexpected peaks or valleys—to encourage reflection. For example, a system could allow individuals to examine relationships between cohort data and personal data by highlighting differences in highs and lows, thus prompting inquiry-based reflection. Viewed in terms of Fleck and Fitzpatrick's [28] levels of reflection and the computing capabilities needed to support them, system designers can enable individuals to *"look for relationships between pieces of experience or knowledge"* and *"consider other points of view."*

5.4 Limitations and Future Work

By observing students as they engaged with real-world *physical activity*, *digital social activity*, and *stress* data, our study provides insights into how the data of "people like me" affect an individual's self-reflection process. However, our study has some limitations. The instrumentation of mobile phones and the software platforms we used—while state of the art—impose limits on the data types that we could collect. For example, some participants lamented the absence of inaccessible IP messaging applications such as WhatsApp. Additionally, while students are an important participant group, they represent only a subset of the population who engage with personal tracking systems.

Our study prototype provided limited exploration capabilities such as simple filtering using a set of check-boxes, and only showed the average of cohort data. In addition, our study used up to three weeks of participants' data from the mobile-sensing study. If provided a richer set of interactions with their data and others, for a longer period of time, people could demonstrate deeper levels of reflection. On the other hand, adding a richer set of interactions could also bring new challenges around usability, interpretability, and learnability.

By supporting visualization of the average of a cohort, we abstracted the collective nature of the social data. For example, average data of all mobile-sensing participants, Everyone, was represented in our visualization as a single line: each Individual participant's data took on the same visual representation as the *aggregated* data of others. It is easy to imagine representing others' data using alternative visualizations, such as a time-series envelope—which visualizes the distribution within a cohort, possibly overlaying the mean or median of the group in a different style. Different visualization choices will impact an individual's reflection-on-action. For example, when an individual is searching for similar graphical patterns our single-line visualization may encourage the individual to more closely identify with a group than if a time-series envelope was used to represent the data. While we believe our choice of visualization to be appropriate for our application, further research should explore how alternative visualizations impact these reflective processes. Discovering what visualizations are most appropriate for different tracked activities remains an open question that is already being discussed in the area of personal informatics systems [53]. Investigating which visualizations are more appropriate for representing aggregate data of others is an important extension to that discussion.

Finally, emerging research on facilitating personal reflection is exploring how to best incorporate a variety of personal contextual data (from personal calendars, detected activities, one's location, local weather, among other data) into reflective presentations, to enable people to make sense of their data. Our study limited the scope of context to focus on the use of others' data. It would be useful to investigate how the data of others—as a type of context—can complement other types of contextual data to facilitate reflection in the future.

Our findings suggest that this space merits further work, especially as ubiquitous computing research explores new ways of leveraging the large repositories of behavioral and physiological data to support personal and social goals. Exploring how reflection unfolds in other settings (e.g., mobile) with different populations and in larger studies will be vital. Doing so will require researchers to attend to a number of privacy considerations associated with enabling mobile reflection (e.g., showing comparative data in-situ during a social activity). Another important area for future work includes addressing the ecological validity of our findings. It will also be important to pursue specific design recommendations for systems that include the data of others in ways that are safe, privacy-preserving, and promote wellness.

6 CONCLUSION

In this paper, we examined how making a large set of anonymous, aggregate data visible alongside individuals' own data—a practice increasingly entering the mainstream—affected the retrospective self-reflection process. Through data review of three types of personal wellness data (*daily physical activity*, *digital social activity*, and *perceived stress*), we investigated how people integrate others' data to make sense of their own data, and how they identify insights and form goals in the absence of pre-existing social ties. Our focus on letting participants freely interact with a diverse set of personal and aggregate wellness data provides insights into how participants use personal informatics systems that make cohort data available. This research serves as a starting point for designers and researchers looking to explore new systems that allow individuals to reflect on data collected from "people like me."

REFERENCES

[1] Jae-eul Bae, Youn-kyung Lim, Jin-bae Bang, and Myung-suk Kim. 2014. Ripening room: designing social media for self-reflection in self-expression. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, Vancouver, BC, Canada, 103–112.

[2] Elizabeth Bales and William Griswold. 2011. Interpersonal informatics: making social influence visible. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, Vancouver, BC, Canada, 2227–2232.

[3] Albert Bandura. 1986. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

[4] Jared S Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F Watson, and Julie Kientz. 2012. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, Austin, TX, USA, 1401–1410.

[5] Eric PS Baumer. 2015. Reflective informatics: conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 585–594.

[6] Eric PS Baumer, Sherri Jean Katz, Jill E Freeman, Phil Adams, Amy L Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M Olson, and Geri K Gay. 2012. Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, New York, NY, USA, 475–484.

[7] Eric PS Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014. Reviewing reflection: on the use of reflection in interactive system design. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, Vancouver, BC, Canada, 93–102.

[8] Nuran Bayram and Nazan Bilgel. 2008. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Social Psychiatry and Psychiatric Epidemiology* 43, 8 (2008), 667–672. https://doi.org/10.1007/s00127-008-0345-x

[9] Larry Chan, Vedant Das Swain, Christina Kelley, Kaya de Barbaro, Gregory D. Abowd, and Lauren Wilcox. 2018. Students' Experiences with Ecological Momentary Assessment Tools to Report on Emotional Well-being. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 3 (March 2018), 20 pages. https://doi.org/10.1145/3191735

[10] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. 2015. SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Umeda, Osaka, Japan, 121–132.

[11] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding Self-Reflection: How People Reflect on Personal Data Through Visual Data Exploration. In *11th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health 2017)*. ACM, Barcelona, Spain.

[12] Eun Kyoung Choe, Nicole B Lee, Bongshin Lee, Wanda Pratt, and Julie A Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1143–1152.

[13] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver, CO, USA, 1674–1687.

[14] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. 1983. A Global Measure of Perceived Stress. *Journal of Health and Social Behavior* 24, 4 (1983), 385–396. https://doi.org/10.2307/2136404

[15] Sheldon Ed Cohen and S1 Syme. 1985. *Social support and health.* Academic Press.

[16] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1797–1806. https://doi.org/10.1145/1357054.1357335

[17] Andrea Cuttone, Michael Kai Petersen, and Jakob Eg Larsen. 2014. Four data visualization heuristics to facilitate reflection in personal informatics. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, Heraklion, Crete, Greece, 541–552.

[18] Angela Lee Duckworth and Patrick D Quinn. 2009. Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment* 91, 2 (2009), 166–174.

[19] Chris Elsden, David S. Kirk, and Abigail C. Durrant. 2016. A Quantified Past: Toward Design for Remembering With Personal Informatics. *Human-Computer Interaction* 31, 6 (2016), 518–557. https://doi.org/10.1080/07370024.2015.1093422

[20] Chris Elsden, Aisling O'Kane, Paul Marshall, Abigail Durrant, Rowanne Fleck, John Rooksby, and Deborah Lupton. 2017. Quantified Data & Social Relationships. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2017), 644–651. https://doi.org/10.1145/3027063.3027065

[21] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, USA, 731–742. https://doi.org/10.1145/2750858.2804250

[22] Anja Exler, Andrea Schankin, Christoph Klebsattel, and Michael Beigl. 2016. A wearable system for mood assessment considering smartphone features and data from mobile ECGs. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, Heidelberg, Germany, 1153–1161.

[23] Tom Fawcett. 2015. Mining the Quantified Self: Personal Knowledge Discovery as a Challenge for Data Science. *Big Data* 3, 4 (2015), 249–266. https://doi.org/10.1089/big.2015.0049

[24] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2015. AWARE: mobile context instrumentation framework. *Frontiers in ICT* 2 (2015), 6.

[25] Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.

[26] Fitbit. 2018. Start Sleeping Better with Fitbit. https://www.fitbit.com/sleep-better

[27] Rowanne Fleck and Geraldine Fitzpatrick. 2009. Teachers' and tutors' social reflection around SenseCam images. *International Journal of Human Computer Studies* 67, 12 (2009), 1024–1036. https://doi.org/10.1016/j.ijhcs.2009.09.004

[28] Rowanne Fleck and Geraldine Fitzpatrick. 2010. 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*. ACM, Brisbane, Australia, 216–223. https://doi.org/10.1145/1952222.1952269

[29] Andrew Garbett, David Chatting, Gerard Wilkinson, Clement Lee, and Ahmed Kharrufa. 2018. ThinkActive: Designing for Pseudonymous Activity Tracking in the Classroom. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 7.

[30] Roland Gasser, Dominique Brodbeck, Markus Degen, Jürg Luthiger, Remo Wyss, and Serge Reichlin. 2006. Persuasiveness of a mobile lifestyle coaching application using social facilitation. *Persuasive technology* (2006), 27–38.

[31] Google. 2017. DetectedActivity. https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity

[32] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How do we engage with activity trackers?: a longitudinal study of Habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Osaka, Japan, 1305–1316. https://doi.org/10.1145/2750858.2804290

[33] Lisa Graham, Anthony Tang, and Carman Neustaedter. 2016. Help Me Help You: Shared Reflection for Personal Data. In *Proceedings of the 19th International Conference on Supporting Group Work*. ACM, Sanibel Island, FL, USA, 99–109.

[34] Weixi Gu, Yunxin Liu, Yuxun Zhou, Zimu Zhou, Costas J Spanos, and Lin Zhang. 2017. BikeSafe: bicycle behavior monitoring via smartphones. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, Maui, HI, USA, 45–48.

[35] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. 2016. Fostering engagement with personal informatics systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. ACM, Brisbane, QLD, Australia, 286–300.

[36] Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity tracking: barriers, workarounds and customisation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Osaka, Japan, 617–621. https://doi.org/10.1145/2750858.2805832

[37] Julie D Henry and John R Crawford. 2005. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. *The British journal of clinical psychology / the British Psychological Society* 44, Pt 2 (2005), 227–239. https://doi.org/10.1348/014466505X29657

[38] Shawaz Iqbal, Sandhya Gupta, and E Venkatarao. 2015. stress , anxiety & depression among medical undergraduate students & their socio-demographic correlates. *INDIAN J MED RES* 141, March (2015), 354–357.

[39] Matthew Kay. 2014. Challenges in personal health tracking: the data isn't enough. *XRDS: Crossroads, The ACM Magazine for Students* 21, 2 (2014), 32–37. https://doi.org/10.1145/2678024

[40] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. 2012. Lullaby: a capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, Pittsburgh, PA, USA, 226–234.

[41] Christina Kelley, Bongshin Lee, and Lauren Wilcox. 2017. Self-tracking for Mental Wellness: Understanding Expert Perspectives and Student Experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver, CO, USA, 629–641.
[42] Danielle Kosecki. 2017. New Fitbit Features Deliver Data Previously Only Available Through a Sleep Lab. https://blog.fitbit.com/sleep-stages-and-sleep-insights-announcement/
[43] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, Beijing, China, 405–414.
[44] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. FishâĂŹnâĂŹSteps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, Berlin, Heidelberg, 261–278. https://doi.org/10.1007/11853565_16
[45] Julie Maitland and Matthew Chalmers. 2011. Designing for Peer Involvement in Weight Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 315–324. https://doi.org/10.1145/1978942.1978988
[46] Andrew D Miller and Elizabeth D Mynatt. 2014. StepStream: a school-based pervasive social fitness system for everyday adolescent health. , 2823–2832 pages.
[47] Amber D Miller, Jessica Pater, and Elizabeth D Mynatt. 2013. Design strategies for youth-focused pervasive social health games. , 9–16 pages.
[48] Sean A. Munson, Hasan Cavusoglu, Larry Frisch, and Sidney Fels. 2013. Sociotechnical challenges and progress in using social media for health. *Journal of Medical Internet Research* 15, 10 (2013), 1–14. https://doi.org/10.2196/jmir.2792
[49] Jeungmin Oh and Uichin Lee. 2015. Exploring UX issues in Quantified Self technologies. In *Mobile Computing and Ubiquitous Networking (ICMU), 2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking, ICMU 2015*. IEEE, Auckland, New Zealand, 53–59. https://doi.org/10.1109/ICMU.2015.7061028
[50] Kunwoo Park, Ingmar Weber, Meeyoung Cha, and Chul Lee. 2016. Persistent sharing of fitness app status on twitter. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 184–194.
[51] Andrea Grimes Parker. 2014. Reflection-through-performance: personal implications of documenting health behaviors for the collective. *Personal and Ubiquitous Computing* 18, 7 (01 Oct 2014), 1737–1752. https://doi.org/10.1007/s00779-014-0780-5
[52] Tamara Peyton, Erika Poole, Madhu Reddy, Jennifer Kraschnewski, and Cynthia Chuang. 2014. Every pregnancy is different: designing mHealth for the pregnancy ecology. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, Vancouver, BC, Canada, 577–586.
[53] Bernd Ploderer, Wolfgang Reitberger, Harri Oinas-Kukkonen, and Julia van Gemert-Pijnen. 2014. Social interaction and reflection for behaviour change.
[54] Aare Puussaar, Adrian K Clear, and Peter Wright. 2017. Enhancing Personal Informatics Through Social Sensemaking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver, CO, USA, 6936–6942. https://doi.org/10.1145/3025453.3025804
[55] Quedget. 2017. Quedget. https://quedget.com
[56] Samuel Reade, Karen Spencer, Jamie C Sergeant, Matthew Sperrin, David M Schultz, John Ainsworth, Rashmi Lakshminarayana, Bruce Hellman, Ben James, John McBeth, et al. 2017. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR mHealth and uHealth* 5, 3 (2017).
[57] Attila Reiss and Didier Stricker. 2013. Personalized mobile physical activity recognition. In *Proceedings of the 2013 international symposium on wearable computers*. ACM, Zurich, Switzerland, 25–28.
[58] Saeyoung Rho, Injung Lee, Hankyung Kim, Jonghyuk Jung, Hyungi Kim, Bong Gwan Jun, and Youn-kyung Lim. 2017. FutureSelf: What Happens When We Forecast Self-Trackers? Future Health Statuses?. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, Edinburgh, United Kingdom, 637–648.
[59] Verónica Rivera-Pelayo, Valentin Zacharias, Lars Müller, and Simone Braun. 2012. Applying quantified self approaches to support reflective learning. In *Proceedings of the 2nd international conference on learning analytics and knowledge*. ACM, Vancouver, BC, Canada, 111–114.
[60] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, Toronto, Canada, 1163–1172. https://doi.org/10.1145/2556288.2557039
[61] Petr Slovák, Christopher Frauenberger, and Geraldine Fitzpatrick. 2017. Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2696–2707. https://doi.org/10.1145/3025453.3025516
[62] Tableau Software. 2018. Tableau Software. https://www.tableau.com.

[63] Tammy Toscos, Anne Faber, Shunying An, and Mona Praful Gandhi. 2006. Chick clique: persuasive technology to motivate teenage girls to exercise. , 1873–1878 pages.

[64] Yonatan Vaizman, Katherine Ellis, Gert Lanckriet, and Nadir Weibel. 2018. Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, MontrÃĺal, Canada, 554.

[65] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Seattle, WA, USA, 3–14.

Received May 2018; revised July 2018; accepted September 2018