Supporting Reliable Data Collection in Self-Experimentation

Sayali Phatak
Arizona State University
Designing Health Lab –
School of Nutrition and Health Promotion
Phoenix, AZ 85004, USA
sayali.phatak@asu.edu

Jisoo Lee
Arizona State University
Designing Health Lab –
School of Nutrition and Health Promotion
Phoenix, AZ 85004, USA
Jisoo.Lee@asu.edu

Marios Hadjimichael
Arizona State University
Designing Health Lab –
School of Nutrition and Health Promotion
Phoenix, AZ 85004, USA
marioshadj@asu.edu

Submitted for CHI'17, May 6 -11, 2017, Denver, CO, USA

Abstract
Self-experimentation has seen a tremendous increase in popularity, but users' current methods often lack the desirable scientific rigor. One of the issues is related to data accuracy. Self-tracking often involves some degree of self-reported data capture. Using established psychometric principles in the design of self-report protocols is crucial to ensure that the data is reliable. It is a complicated process that is largely ignored when users create their own measures. But to date, there are no tools that help users in this process. We propose a knowledge-based recommender system that aims to guide users in designing their data collection tool using existing domain knowledge, psychometric design heuristics, and what it knows about the user's research questions and relevant contextual information. We initially focus on building such a system for sleep-related manual data collection.

Author Keywords
self-experimentation; self-tracking; co-design

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Experimentation, Reliability

Introduction
Self-experimentation has been increasing in popularity, but has many issues associated with it that bring into

User-specified Constraints

Information users provide about their research question and tracking preferences. Examples:

- **Research Question/Purpose**
  e.g., *Does alcohol affect my sleep quality?*
- **Demographic Information**
  e.g., gender, age
- **Desired tracking method**
  e.g., pen & paper, Google form
- **Desired tracking frequency**
  e.g., once/day
- **Desired number of items to track**
- **Time investment**
  e.g., up to 5 minutes/day

Knowledge-base

- **Database of pre-designed self-report protocols** that measure:
  - **Sleep-related behaviors**
    e.g., sleep-quality
  - **Contextual factors**
    e.g., ambient light, caffeine

- **Protocol attributes**
  Examples for a protocol measuring sleep quality:
  - **Nature of phenomenon**
    e.g., discrete (event-based)
  - **Number of items in protocol**
    3 (sleep quality, number of awakenings, rating how refreshing sleep was)
  - **Recommended frequency of measurement**
    e.g. once/day, within an hour of waking up
  - **Recommended sampling strategy**
    e.g., time-based
  - **Type of answer formats**
    e.g Likert scale
  - **Completion time**
    e.g., ~1 minute

Table 1a. Aspects of a knowledge-based recommender system [1]

question the validity of such experiments. While some of the problems are difficult to avoid (e.g., researcher bias, reactivity), many of them can often be minimized by sound data collection. Poor design of data collection protocols is one of the most common problems associated with self-experimentation. The scientific community has made tremendous progress in the design and analysis of single-case (n=1) experiments, but this work hinges on robust measurements. We need to develop platforms to share such knowledge and better assist the self-experimentation community in their research endeavors, but in a way that does not hinder creativity.

Barriers associated with data-collection

Previous work has identified common barriers related to data collection in the Quantified Self community. Choe et al., in their study looking at the self-tracking practices of quantified-selfers noted that participants focused too much on tracking 'symptoms' but forgot to track contextual factors and triggers [5]. Another study also noted that users often did not have the necessary contextual information to be able to reflect on the data [8]. It has also been observed that insufficient support in deciding what to track often leads to over-tracking, making this process highly burdensome. A proactive and decision-driven approach might be more feasible and low-burden than collecting a multitude of information and retrospectively asking questions about the collected data.

Design of self-reported data collection protocols is complicated.

Most self-experimentation endeavors involve some degree of self-reported data capture. Self-reporting is very important when the focus is on understanding the user's subjective states and experiences. Additionally, there are certain items that cannot be effectively tracked in a passive manner [3,7]. Automated data collection might also reduce awareness and self-reflection that is more likely to occur when the user is involved in the data collection process [5,7]. On the flip side, measuring subjective states has unique challenges associated with it that require careful thinking in order to minimize error.

Ecological Momentary Assessment (EMA) or Experience Sampling are methods that emphasize the repeated capture of data in real time to assess human experiences, behavior and states, thus reducing recall bias [14]. Their design requires heuristics and principles from the field of psychometrics, which is the discipline concerned with evaluating the attributes of psychological and behavioral measures [6]. A robust measure is one that is **accurate,** it measures the construct (e.g., daytime sleepiness) that it is intended to measure, **reliable**, meaning it is a stable and consistent measure of the construct, and is **sensitive** to capturing change, meaning it has the ability to detect meaningful differences in the construct [6]. There are numerous attributes of the measures that affect these properties for a particular measure, such as the type of scale used (e.g., Likert vs. semantic differential), number of points on scale (Yes/No vs. 5-point scale), framing (e.g., ensuring that you operationally define the construct you are interested in, being specific).

In addition to the measures, the other aspects of data collection (contingencies, measurement frequency, duration) should also ideally be based on attributes of the phenomenon being measured, such as the type (discrete, e.g., alcohol intake vs continuous, e.g., stress), natural fluctuations over time (e.g., mood fluctuates frequently throughout the day while an individual's height remains more or less the same) and the minimum duration of measurement for it to be a reliable assessment of the phenomenon.

It is rare that self-experimenters are trained in measurement development and yet often create their own measures using tools like Excel and Google forms [5]. Nuanced design decisions tend to get overlooked and measures are often created with little regard for

Decision Rules

Set of rules to match user's research question and preferences to suitable protocols

Recommendation

Examples:

- **Pre-designed protocol and explanation**
  Example formats: PACO, Google form, PDF
- **Important information related to construct(s) under question**
- **Reasoning behind recommendation**

Table 1b. Aspects of a knowledge-based recommender system [1]

psychometric quality. This is unsurprising, seeing that this process is time-consuming and burdensome, but at the same time, reliable data collection is paramount to making causal inferences from collected data. **There is a great opportunity for tools that can aid individuals in developing more decision-driven and robust data collection protocols.**

Current state of self-tracking tools

Behavior-specific commercial tools support self-tracking, but often offer no support when users have more specific or diagnostic questions about their behavior [9]. Even when they do, the reliability of self-report measures is debatable, since they often simplify constructs that are multidimensional for the purposes of lowering user-burden [4,12; 9] On the other hand, experience sampling tools such as the Personal Analytics COmpanion (PACO) [10] offer the flexibility and functionality to let users create their own measurement protocols, but without any research question-specific support. Websites such as the Quantified Self [QS; 11] have numerous posts providing guidance, but unfortunately it is often too unstructured to be useful for novice users. Users are either restricted to tracking goals that commercial apps allow or have to resort to creating their own measures.

Based on the above review of the literature, we identified key issues in this area:

- Sound psychometrics are important for robust data collection, but often lacking.
- We need decision-support tools that balance the needs of the user (e.g., flexibility to ask questions; number of items tracked to reduce burden) with the desire to develop robust, decision-driven protocols.

Knowledge-based decision support for tracking sleep behavior

For the initial design of such a system, we focus on sleep-related manual tracking. Sleep tracking is one of the most common behaviors that are self-tracked [4,5]. In addition, it has also been suggested that both, the subjective and objective aspects of sleep are important and should be considered separately, establishing the need for manual tracking in this area [2]. Moreover, a wide range of contextual factors influence sleep behavior, and they're idiosyncratic in how they affect different people [4,9]. While many sleep-tracking apps exist, and some of them even offer trigger tracking [12], they are restrictive in terms of the type of data users can collect.

In order to balance the needs of the end-users and of the research process, we can build platforms that assist users in their self-experimentation endeavors by guiding them through the protocol design process. Suitable existent protocols can also be recommended using advanced experience sampling tools such as PACO or using simpler tools such as Google Forms, both of which have the flexibility of being modified if the user wants, to support more iterative learning.

One approach to solve this problem is developing a knowledge-based recommender system (e.g., www.hotpads.com, which is an apartment search website). Such systems use knowledge about users and the products to engage in a knowledge-based approach to generating a recommendation [1,13]. The interface can also be designed to be conversational, to walk the user through the decision process. The use of such systems can be helpful when only a few individuals have a majority of the knowledge and to distribute knowledge in a more cost-effective manner. They also allow users to specify constraints or hard requirements on the recommended items [1]. Potential aspects of such a system in a behavioral context are displayed in Table 1a and 1b.

Research Agenda

Agenda involves formative work with key stakeholders: novice and extreme self-experimenters, and EMA/sleep experts to inform the design of all the components of

Research Questions:

1. What is the current process that self-trackers use when creating their own self-report tools?
2. How might we support this process in a way that fits with their current approach?
3. What are the important protocol attributes that are vital to achieve a desirable level of scientific rigor in sleep-related self-experimentation?

Table 2. Research Questions to inform initial formative work.

such a system. Few of the initial research questions are presented in Table 2.

Self-experimentation design workshop

In the workshop, we will share insights from some of our formative work along with initial prototypes. We would like to use the hands-on activities in the workshop to invite discussion about the different approaches that can be used to design such a system, potential limitations and other design considerations. Some ideas for hands-on activities to dive into the process of protocol development include:

1. Self-tracking tool creation activity: Participants create their own measures using for specific research questions to experience the process.
2. Mock-up of a support system using pairs of domain experts and non-experts: The expert guides the non-expert in developing a self-tracking protocol.
3. Building prototypes of an interactive website that guides users in designing their self-tracking measures.

References

1. Aggarwal, C. C. (2016). *Recommender Systems*.
2. Baker, F. C., Maloney, S., & Driver, H. S. (1999). A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women. *Journal of psychosomatic research*, *47*(4), 335-341.
3. Cena, F., Lombardi, I., Rapp, A., & Sarzotti, F. (2014). Self-Monitoring of Emotions: a Novel Personal Informatics Solution for an Enhanced Self-Reporting. In *UMAP Workshops*.
4. Choe, E. K., Lee, B., Kay, M., Pratt, W., & Kientz, J. A. (2015, September). 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* (pp. 121-132). ACM.
5. Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014, April). Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1143-1152). ACM.
6. Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage.
7. Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. *Proceedings of the 28th International Conference on Human Factors in Computing Systems CHI 10*, (November 2015), 557. http://doi.org/10.1145/1753326.1753409
8. Li, I., Dey, A., & Forlizzi, J. (2011). Understanding My Data, Myself: Supporting Self-Reflection with Ubicomp Technologies. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 405–414. http://doi.org/10.1145/2030112.2030166
9. Liang, Z., Ploderer, B., Liu, W., Nagata, Y., Bailey, J., Kulik, L., & Li, Y. (2016). SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Personal and Ubiquitous Computing*, *20*(6), 985-1000.
10. PACO: The Personal Analytics Companion. https://www.pacoapp.com/
11. Quantified Self: http://quantifiedself.com/
12. Runtastic GmBH. Sleep Better. https://www.runtastic.com/en/apps/sleepbetter. Accessed on Feb 22, 2017.
13. Sajja, P. S., & Akerkar, R. (2010). Knowledge-based systems for development. *Advanced knowledge based systems: Model, Application & Research*, 1-11.
14. Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*.