Opportunities and Challenges for Self-Experimentation in Self-Tracking

Ravi Karkar
Computer Science & Engineering
DUB Group, University of Washington
Seattle, WA 98195, USA
rkarkar@cs.washington.edu

James Fogarty
Computer Science & Engineering
DUB Group, University of Washington
Seattle, WA 98195, USA
jfogarty@cs.washington.edu

Julie A. Kientz
Human Centered Design & Engineering
DUB Group, University of Washington
Seattle, WA 98195, USA
jkientz@uw.edu

Sean A. Munson
Human Centered Design & Engineering
DUB Group, University of Washington
Seattle, WA 98195, USA
smunson@uw.edu

Roger Vilardaga
Psychiatry and Behavioral Sciences
University of Washington
Seattle, WA 98195, USA
vilardag@uw.edu

Jasmine Zia
Division of Gastroenterology, Medicine
University of Washington
Seattle, WA 98195, USA
jzia@medicine.washington.edu

Ubicomp/ISWC'15 Adjunct , September 07 - 11, 2015, Osaka, Japan

ACM 978-1-4503-3575-1/15/09 $15.00
DOI: http://dx.doi.org/10.1145/2800835.2800949

Abstract
Personal informatics applications support capture and access of data related to an increasing variety of dimensions of everyday life. However, such applications often fail to effectively support diagnostic self-tracking, wherein people seek to answer a specific question about themselves. Current approaches are therefore difficult, tedious, and error-prone. This workshop paper discusses our ongoing efforts to develop methods for self-experimentation in self-tracking. We examine how self-experimentation situates within existing models of personal informatics processes, discuss our current focus on personal food triggers in patients suffering from Irritable Bowel Syndrome, and highlight open challenges for self-experimentation more broadly.

Author Keywords
Self-experimentation; self-tracking; personal informatics.

Introduction
Current self-tracking applications support capture and access of data related to various dimensions of life, including physical activity, food, and sleep. However, most applications still fall short in providing meaningful and actionable feedback. Self-trackers are often provided with graphs of raw data, but left on their own to interpret and decide how to act upon that data [4].

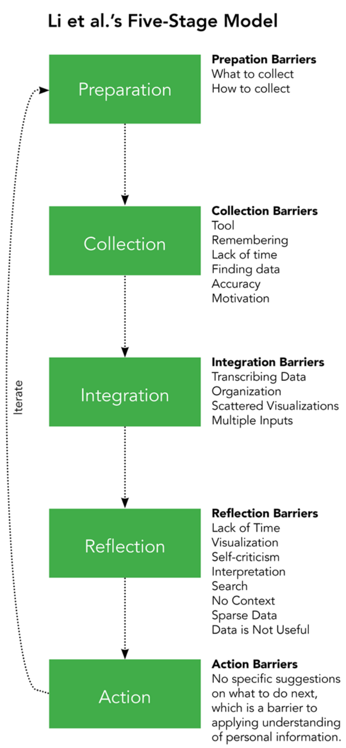

Figure 1: Li et al. present a five-stage model of personal informatics, with barriers that must be overcome in each stage. Self-experimentation can be seen as scaffolding a process of personal scientific discovery within this larger process.

Choe et al. examine the Quantified Self community as an example of "extreme users" of self-tracking [1]. They find that many self-trackers are collecting data to answer specific questions, but either: (1) fail to obtain the desired answer, or (2) obtain the answer only after developing ad hoc solutions for their personal needs (i.e., custom software). Choe et al. further identify three common pitfalls: (1) tracking too much, which quickly leads to burnout and abandonment (2) not tracking necessary triggers and context, thus preventing people from obtaining an answer, and (3) reaching dubious conclusions that lack scientific rigor.

As self-tracking continues to expand into mainstream practice, it becomes increasingly important to support people in using self-tracking data to answer personal questions. Current platforms and applications generally fail to address these pitfalls, so obtaining high-quality answers remains a tedious and error-prone endeavor. We see this as an opportunity for designers to create new tools that better support self-tracker goals.

In this workshop submission, we propose support for self-experimentation within self-tracking, discussing our current work and the broader opportunity. We begin with an introduction to self-experimentation, considering how it aligns with existing models of personal informatics. We then discuss how we are currently incorporating self-experimentation in an application we are developing for patients with Irritable Bowel Syndrome. Finally, we conclude with a discussion of open challenges in employing self-experimentation in other domains of self-tracking. We believe our perspective can contribute a valuable perspective and promote thoughtful discussion in the UbiComp 2015 workshop on New Frontiers of Quantified Self.

Self-Experimentation in Self-Tracking

We model self-experimentation in self-tracking as a three-step process: (1) formulating a hypothesis, (2) testing a hypothesis, and (3) interpreting a result. These can be repeated to test multiple hypotheses. Importantly, they are also intended to be situated within a model of the larger personal informatics process.

For example, we can situate our process in Li et al.'s five-stage model of personal informatics, consisting of Preparation, Collection, Integration, Reflection, and Action [11]. Each stage presents barriers that must be overcome, with failures in early stages cascading into later stages. Preparation requires deciding what and how to collect, collection requires completeness and accuracy, integration requires organizing collected data, reflection requires interpreting and making sense of data, and action requires determining how to convert new understanding into an actionable plan (Fig. 1).

Explicit support for self-experimentation can be seen as re-casting self-tracking as a process of personal scientific discovery, with the goal of helping people successfully navigate these barriers. We aim to scaffold preparation, collection, and integration around the design and execution of self-experiments while supporting reflection and action through interpretation of self-experiment results. Self-experimentation thus aligns itself with design recommendations in personal informatics by providing concrete steps a person can take to navigate these barriers. We thus aim to both (1) reduce burdens and uncertainty in self-tracking, and (2) provide greater certainty and value in results.

Self-experimentation also fits well within other models of personal informatics. Choe et al. recommend providing early guidance in what to track, support for self-experimentation, and balancing the burden of collection against reflection enabled by collection [1]. Rooksby et al.'s work on lived informatics supports viewing data in context and accounting for non-rational aspects of interaction in self-tracking [15]. Epstein et al. propose a model for lived informatics that raises forgetting, upkeep, skipping, and suspending as additional barriers to consider [5].

Self-Experimentation Design and Analysis

We are designing self-experimentation based on single case designs (SCDs), sometimes referred to as n-of-1 studies [12]. SCDs have an individual serve as their own control, therefore making the design sensitive to individual differences. Two or more experimental conditions are defined (e.g., a baseline and any manipulations). Outcome measures are then monitored as these conditions are applied, with the goal of identifying any relationships between the experimental conditions and the outcome measures.

SCDs traditionally suffer from limitations regarding internal validity, a lack of statistical inference, and a limited number of observations [7]. We overcome these by applying randomization tests, wherein a random assignment is given to a population of occasions rather than a population of individuals [2,8]. In contrast to traditional group design where each individual is assigned to a condition, SCDs with randomization tests assign each measurement occasion to a condition. A permutation procedure can then be used to create all possible combinations of condition exposures and outcome measures, then a p value indicating the probability of the null hypothesis. This procedure was originally envisioned by Fisher in the 1930s, ensures the same internal validity as group experiments, eliminates the statistical assumptions of parametric tests, and can therefore be used as a statistical test with both individuals and groups of individuals [6]. The primary limitation remains that it does not provide external validity, but this limitation is not applicable in our case because we use it to personalize known group-based or clinical guidelines to specific individuals. Furthermore, advances in mobile technology now allow more frequent and ecologically valid measurements [16]. Thus, SCDs with randomization tests overcome many limitations of traditional SCDs.

Self-Experimentation in IBS

We are currently incorporating self-experimentation in an application we are developing for patients with Irritable Bowel Syndrome (IBS). IBS is a chronic functional disorder characterized by episodic abdominal pain associated with diarrhea and/or constipation despite normal blood tests, x-rays, and colonoscopies. It affects up to 20% of the US population and is one of the top ten reasons people seek primary care [3,13]. People with IBS report a lower quality of life and consume 50% more healthcare resources than non-IBS counterparts [10,14]. There are many potential triggers for IBS symptom flare-ups: certain foods, eating behaviors, stress, sleep disturbances, and menstruation. It provides an ideal domain for self-experimentation, as many patients learn their potential triggers through trial and error or through more extreme interventions (e.g., elimination diets).

We are designing and developing a mobile application that guides IBS patients through self-experiments they choose in their efforts to identify personal triggers. We are initially focused on food triggers, as they are both

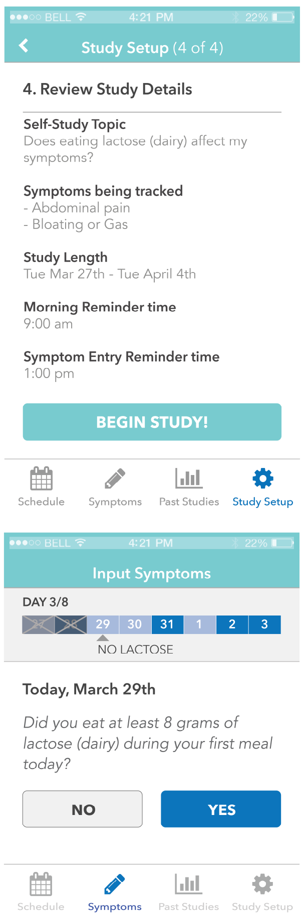

Figure 2: Our approach to self-experimentation in IBS begins by defining a specific symptom and trigger to test (top). This provides context for data collection and reduces burden by narrowing the scope of data to collect (bottom).

common and actionable. This section briefly reviews our ongoing work to support self-experimentation in IBS, discusses a self-experimentation process (Fig. 3), and considers our design approaches to the challenges in outlined in Li et al.'s five-stage model [11].

Formulating Hypotheses

Two major barriers in the preparation stage are determining what data to collect and how to collect it. We address these in formulating a hypothesis.

A person begins a self-experiment by formulating a hypothesis they want to test. This reduces ambiguity by defining a specific symptom and probable trigger to test (i.e., a dependent and independent variable) (Fig. 2). Decisions regarding which symptom and trigger to test can be based in prior personal experience, in population-level understandings of likely triggers, through engagement with medical providers, through peer engagement with other self-trackers, or using other sources of potential hypotheses. Defining a specific testable hypothesis helps a person focus on the context in which data will be collected and reduces burden by narrowing the scope of data to collect.

Testing Hypotheses

Barriers in the collection and integration stages include collecting and organizing appropriate data. We scaffold this process around gathering data to test a hypothesis.

Given a hypothesis to be tested, our application automates experiment design and suggests a study plan with a customized schedule based on a person's selection of study duration and when to start. The mobile application platform can support execution of the experiment by automatically tracking the

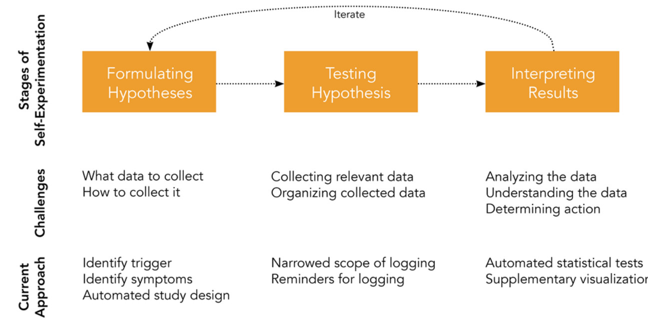

Figure 3: Stages of self-experimentation, challenges to be addressed in each, and our current approaches.

experiment schedule, providing reminders to log symptom and adherence to experimental condition, and by providing in-situ educational materials to help in understanding the experimental process or how to properly adhere to the current food trigger experiment.

Interpreting Results

Barriers in the reflection and action stages include analyzing and understanding the data as well as determining concrete actions that can be taken.

At the end of a self-experiment, appropriate statistical tests are run and a p value is reported (Fig. 4). A visualization is also provided to aid in reflection. SCDs are often analyzed visually to look for trends and infer relationships, but a randomized SCD can make it challenging to visually analyze trends. Conditions can be distributed in any order, with no fixed phase lengths, making it difficult to find patterns across phases. To overcome this we designed a visualization inspired by a violin plot [9]. It removes temporal information and instead focuses on the distribution of the outcome measure across different conditions.

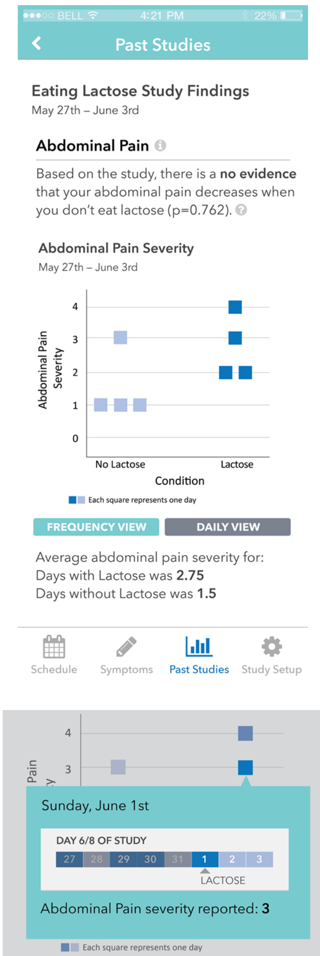

Figure 4: Results of a self-experiment are presented as a statistical test with a visualization of results and support for drilling into individual entries.

Drilling into individual entries can allow viewing any personal notes or photographs attached to measures. Based on the results of the statistical analysis and a person's experiences in an experiment, they can decide whether the results suggest and warrant removing or controlling a potential trigger food.

Self-Experimentation Opportunities

We have focused on self-experimentation in IBS, but our work suggests self-experimentation opportunities in other self-tracking domains. We have identified several challenges that need to be considered in generalizing our current approaches to self-experimentation.

Temporal Relationships

Our current approach relies upon symptoms following trigger foods after a short period of time (i.e., hours). It also relies on measurements being independent after a day has passed (i.e., the body "resets" overnight). If either of these are not true in another domain, our approach will need to be modified to account for this.

Difficulty in Defining and Manipulating a Variable

Food triggers are relatively easy to define and experimentally manipulate, but additional challenges are presented in other self-experimentation domains. For example, stress can serve as a trigger for IBS and can also impact sleep. But it is less obvious how to operationalize or experimentally manipulate stress. Additional challenges will also arise when expanding our work to consider multiple independent variables.

Drift and Biases in Self-Reported Measures

Self-reported measures can be a concern for internal validity, but are likely inherent to many applications of self-experimentation. Self-reports may drift over time and may be influenced by other factors (e.g., mood, stress, life events). Randomization mitigates this, but designers of self-experimentation applications need to be mindful of these and other potential confounds.

Knowledge of Relevant Self-Experiments

Our team includes a gastroenterologist, and so our current work is informed by an appropriate medical perspective on design parameters (e.g., common food triggers, appropriate windows for reporting symptoms, educational material). Although it may in theory be possible to support arbitrary experiments, integrating such domain knowledge into self-experimentation applications can improve the quality of available self-experiments and reduce burdens associated with failed or fundamentally flawed self-experiments.

Lived Informatics and Self-Experimentation

Self-experimentation occurs in a context of daily life. Future deployments of need to explore how these approaches are impacted by such factors as travel, a phone battery dying, or other circumstances related to either an experimental manipulation or an outcome.

Conclusion

Self-tracking and personal informatics continue to grow and move into the mainstream. Analyses of current practices have found support for self-experimentation is lacking, even for "expert users". We believe there are important opportunities for new self-experimentation methods, applications using those methods, and tools supporting developers of such applications.

We have focused this workshop paper on characterizing self-experimentation relative to Li et al.'s five-stage model of personal informatics, describing our ongoing work in IBS, and a non-exhaustive set of challenges that researchers and designers might face in generalizing our approach. Not all self-tracking problems are appropriate for self-experimentation. But where appropriate, our approach to scaffolding the personal informatics process can potentially both reduce the challenges of self-tracking and improve the value obtained. We overall believe this approach warrants further examination in a variety of domains.

We look forward to participation as an opportunity to discuss self-experimentation, its potential benefits, its potential challenges, and opportunities for tool support. The UbiComp 2015 workshop on New Frontiers of Quantified Self is ideal for such a conversation.

Acknowledgements

This research has been funded in part by the University of Washington Innovation Award, the Intel Science and Technology Center for Pervasive Computing, Nokia Research, the National Science Foundation under awards OAI-1028195 and SCH-1344613, the Agency for Healthcare Research Quality under award 1R21HS023654, and the National Institute on Drug Abuse under award 1K99DA037276-01.

References

1. Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. *CHI 2014*, 1143-1152.
2. Eugene S. Edgington, and Patrick Onghena. (2007). *Randomization Tests*.
3. Sigrid Elsenbruch. (2011). Abdominal Pain in Irritable Bowel Syndrome: A Review of Putative Psychological, Neural and Neuro-Immune Mechanisms. *Brain, Behavior, and Immunity*, 25(3), 386-394
4. Daniel A. Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean A. Munson. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. *DIS 2014*, 667-676.
5. Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. A Lived Informatics Model of Personal Informatics. *Ubicomp 2015,* to appear.
6. Ronald A. Fisher. (1935). *The Design of Experiments*.
7. Steven C. Hayes. (1981). Single Case Experimental Design and Empirical Clinical Practice. *Journal of Consulting and Clinical Psychology*, 49(2), 193-211.
8. Mieke Heyvaert, and Patrick Onghena. (2014). Randomization Tests for Single-Case Experiments: State of the Art, State of the Science, and State of the Application. *Journal of Contextual Behavioral Science*, 3(1), 51-64.
9. Jerry L. Hintze, and Ray D. Nelson. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2), 181-184.
10. Uri Ladabaum, Erin Boyd, Wei K. Zhao, Ajitha Mannalithara, Annie Sharabidze, Gurkirpal Singh, Elaine Chung, and Theodore R. Levin. (2012). Diagnosis, Comorbidities, and Management of Irritable Bowel Syndrome in Patients in a Large Health Maintenance Organization. *Clinical Gastroenterology and Hepatology*, 10(1), 37-45.
11. Ian Li, Anind Dey, and Jodi Forlizzi. A Stage-Based Model of Personal Informatics Systems. *CHI 2010*, 557-566
12. Elizabeth O. Lillie, Bradley Patay, Joel Diamant, Brian Issell, Eric J. Topol, and Nicholas J. Schork. (2011). The N-of-1 Clinical Trial: The Ultimate Strategy for Individualizing Medicine? *Personalized Medicine*, 8(2), 161-173.
13. Rebecca M. Lovell, and Alexander C. Ford. (2012). Effect of Gender on Prevalence of Irritable Bowel Syndrome in the Community: Systematic Review and Meta-Analysis. *American Journal of Gastroenterology*, 107(7), 991-1000.
14. Debanjali Mitra, Keith L. Davis, and Robert W. Baran. (2011). All-Cause Health Care Charges Among Managed Care Patients with Constipation and Comorbid Irritable Bowel Syndrome. *Postgraduate Medicine*, 123(3), 122-132.
15. John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal Tracking as Lived Informatics. *CHI 2014*, 1163-1172.
16. Roger Vilardaga, Jonathan B. Bricker, and Michael G. McDonell. (2014). The Promise of Mobile Technologies and Single Case Designs for the Study of Individuals in their Natural Environment. *Journal of Contextual Behavioral Science*, 3(2), 148-153.